



Numéro National de Thèse : 2017LYSEN075

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée par
l'Ecole Normale Supérieure de Lyon

Ecole Doctorale N° 512
Ecole Doctorale en Informatique et Mathématiques de Lyon

Spécialité de doctorat : Informatique

Soutenue publiquement le 29/09/2017, par :
Matteo MORINI

Tools for Understanding the Dynamics of Social Networks

Des Outils pour Comprendre les Dynamiques des Réseaux Sociaux

---

Devant le jury composé de :

Guillaume, Jean-Loup	Professeur	Universté de La Rochelle	Rapporteur
Chavalarias, David	Dir. Rech.	EHESS	Rapporteur
Gauvin, Laetitia	Princ. Rech.	ISI	Examinatrice
Magnien, Clémence	Princ. Rech.	ISI	Examinatrice
Roth, Camille	Assoc. Prof.	Sciences Po, Médialab	Examinateur
Fleury, Eric	Professeur	ENSL, INRIA, LIP	Directeur de thèse
Karsai, Màrton	MDC	ENSL, INRIA, LIP	Co-encadrant
Jensen, Pablo	Professeur	ENSL, CNRS, LP	Co-encadrant

ECOLE NORMALE SUPÉRIEURE DE LYON

DOCTORAL THESIS

Tools for Understanding the Dynamics of Social Networks

Author:
Matteo MORINI

Director:
Prof. Eric FLEURY

Supervisors:
Prof. Pablo JENSEN
Dr. Márton KARSAI

DANTE
Institut Rhônalpin des Systèmes Complexes
Laboratoire de l'Informatique du Parallélisme

ENS DE LYON

September 2017

"Simplicity is a great virtue but it requires hard work to achieve it and education to appreciate it. And to make matters worse: complexity sells better."

Edsger W. Dijkstra

Ecole Normale Supérieure de Lyon

Abstract

Laboratoire de l'Informatique du Parallélisme

Institut Rhônalpin des Systèmes Complexes

Doctor of Philosophy

Tools for Understanding

the Dynamics of Social Networks

by Matteo MORINI

This thesis provides the reader with a compendium of applications of network theory; tailor-made tools suited for the purpose have been devised and implemented in a data-driven fashion.

In the first part, a novel centrality metric, aptly named "bridgeness", is presented, based on a decomposition of the standard betweenness centrality. One component, local connectivity, roughly corresponding to the degree of a node, is set apart from the other, which evaluates longer-range structural properties. Indeed, the latter provides a measure of the relevance of each node in "bridging" weakly connected parts of a network; a prominent feature of the metric is its agnosticism with regard to the eventual ground truth community structure.

A second application is aimed at describing dynamic features of temporal graphs which are apparent at the mesoscopic level. The dataset of choice includes 40 years of selected scientific publications. The appearance and evolution in time of a specific field of study ("wavelets") is captured, discriminating persistent features from transient artifacts, which result from the intrinsically noisy community detection process, independently performed on

successive static snapshots. The concept of “laminar stream”, on which the “complexity score” we seek to optimize is based, is introduced. In a similar vein, a network of Japanese investors has been constructed, based on a dataset which includes (indirect) information on co-owned overseas subsidiaries. A hotly debated issue in the field of industrial economics, the Miwa-Ramseyer hypothesis, has been conclusively shown to be false, at least in its weak form.

Résumé en français

Cette thèse fournit au lecteur un recueil d'applications de la théorie des graphes ; à ce but, des outils sur mesure, adaptés aux applications considerées, ont été conçus et mis en œuvre de manière inspirée par les données.

Dans la première partie, une nouvelle métrique de centralité, nommée "bridgeness", est présentée, basée sur une décomposition de la centralité intermédiaire ("betweenness centrality") standard. Une composante, la "connectivité locale", correspondante approximativement au degré d'un noeud, est différenciée de l'autre, qui, en revanche, évalue les propriétés structurelles à longue distance. En effet, cette dernière fournit une mesure de l'efficacité de chaque noeud à "rélayer" parties faiblement connectées d'un réseau ; une caractéristique importante de cette métrique est son agnosticisme en ce qui concerne la structure de la communauté sous jacente éventuelle.

Une deuxième application vise à décrire les caractéristiques dynamiques des graphes temporels qui apparaissent au niveau mésoscopique. L'ensemble de données de choix comprend 40 ans de publications scientifiques sélectionnées. L'apparition et l'évolution dans le temps d'un domaine d'étude spécifique (les ondelettes) sont capturées, en discriminant les caractéristiques persistantes des artefacts transitoires résultants du processus de détection des communautés, intrinsèquement bruité, effectué indépendemment sur des instantanées statiques successives. La notion de "flux laminaire", sur laquelle repose le "score de complexité" que nous cherchons à optimiser, est présentée.

Dans le même ordre d'idées, un réseau d'investisseurs japonais a été construit, sur la base d'un ensemble de données qui comprend des informations (indirectes) sur les filiales étrangères en copropriété. Une question très débattue dans le domaine de l'économie industrielle, l'hypothèse de Miwa-Ramseyer, a été démontrée de manière concluante comme fausse, du moins sous sa forme forte.

Contents

Abstract **v**

Introduction **1**

1 Detecting global bridges in networks **7**

1.1 Introduction 7

1.2 Measuring bridgeness 9

1.3 Computing global bridges from a community structure 12

1.4 Synthetic network: unbiased LFR 13

1.5 Real network 1: airport's network 15

1.6 Real network 2: scientometric network of ENS Lyon 18

1.7 Discussion 19

2 Japanese business networks in ASEAN countries **23**

2.1 Introduction 23

2.2 Non-Japanese and Japanese business networks 27

2.3 Dataset and descriptive statistics 30

2.3.1 Worldwide waves of investment 33

2.3.2 Heterogeneity in ownership and prevalence of manufacturing 33

2.3.3 Heterogeneity in subsidiaries size and capitalization . 36

2.3.4 Co-investments 37

2.4 Hypotheses 38

2.5 Methodology 41

2.5.1 Community detection on inferred business network . . 41
2.5.2 Independence test 43
2.5.3 Business ties as weighted links 44
2.6 Results and discussion 45
2.7 Conclusions 46

3 Revealing evolutions in dynamical networks 49
3.1 Introduction 49
3.2 Method description 50
3.3 Emergence and evolution of a new scientific field: wavelet analysis 54
3.4 Automatic wavelets history 56
3.5 Test of the method 58
3.6 Comparison to other methods 59
3.7 Concluding remarks and future work 60

A Supplementary Information 61
A.1 Modified Brandes algorithm 61
A.2 Case study on a synthetic network community 62
A.3 Japanese co-investment network facts 63
A.3.1 A small digression on Bridgeness Centrality 65
A.4 Geographical analysis 66
A.4.1 Northern America (incl. U.S. and Canada) 66
A.4.2 ASEAN Countries 67
A.4.3 China 69
A.5 List of WLTC selected authors 70
A.6 Re-splitting unduly merged communities 70
A.7 Network structure examples, colors according to community . 71

Bibliography 75

List of Figures

1.1 Simple graph: BC vs. Bridgeness 11
1.2 LFS synthetic network 14
1.3 Betweenness Centrality vs. Bridgeness 16
1.4 Airports comparison 17
1.5 Bibliometric network 20

2.1 Japanese co-investment overseas 34
2.2 % of Japanese FDI in USD 34
2.3 Heterogeneous network 42
2.4 Joint frequencies of Japanese investors 44
2.5 Japanese investors network 47

3.1 Sliding temporal windows 51
3.2 Ephemeral (artificial) versus structural (real) events 53
3.3 Overview of the history of wavelets 55
3.4 Detail of 1994 split 59
3.5 Modularity evolution, pre- and post-filtering. 59

S1 Zoom on synthetic network 63
S2 Northern America (incl. U.S. and Canada) network 66
S3 ASEAN Network 68
S4 Year 1994 ambiguous nodes 72
S5 Full articles network, time window 1994–1997 73
S6 Full articles network, up to year 1982 74

List of Tables

2.1 Frequencies of firms by country 35
2.2 Employees and capitalization, by macroarea 36
2.3 Employees and capitalization, by industrial sector 36
2.4 Co-investors, by geographical area 37
2.5 Co-investors, by industrial sector 37
2.6 Hypotheses tests results 46

S1 Nodes ranked by decreasing BC 64

Introduction

The unifying thread of this thesis is the approach taken towards a few, seemingly unrelated, problems. On the one hand, data is what drives this research; on the other hand, the tool of choice to gain insight of such input is network analysis.

The *fil rouge* of this work is the elusive concept of communities, which appears in different incarnations throughout the thesis, tailored to suit the peculiarities of the specific problems we tackled.

Advances in this sense are presented for both the graph theoretical and applied empirical approaches, often in a combined fashion.

It may be worth spending a few words on the "behind the scenes" events which happened all along the study process, normally invisible in the final draft of a piece of research. False starts and mishaps are sometimes an object of interest *per se*, providing a different perspective and shedding a different light on the final, polished results.

Our first angle of attack with respect to the "communities issue" has been devising a robust centrality metric capable to capture the concept of "bridging", initially for nodes, and conveniently extended to edges. This process went through several incremental steps. Building upon the standard betweenness centrality (BC) was a matter of common sense, both for its wide diffusion, its well-known properties and the existence of a fast algorithm, capable to scale up to sizeable networks, the kind used in real-world social applications, and going beyond small toy models. The need to untangle the

local and the global components of any given centrality metric was immediately obvious, but before taking its final form, a series of experiments were carried on with a slightly different - and intuitively promising - definition. In particular, the shortest paths (σ) computed, per the definition of BC, between any two nodes, were originally differentiated between long- and short-range, according to a tunable parameter whose value could be sensibly adjusted to any integer value $Pl \geq 3$. According to this intuition, later discarded, only σ at least Pl long were to be included in the sum at the numerator of the standard $BC(j) = \sum_{i \neq j \neq k} \frac{\sigma_{ik}(j)}{\sigma_{ik}}$ formula, where every shortest path σ going from node i to k through j is taken into account, regardless of its length.

This alternate implementation (`MSP`, Minimum Shortest Path) can be tested against the published one (`EXN`, EXcluding Neighbors) in the Gephi plugin published at: `https://github.com/mmorini/gephi-plugins`. Similar results were attained, although at the cost of additional complexity: the "minimum Pl length" being an unwanted degree of freedom, it needs to be tuned ad-hoc according to the characteristics of the graph considered; what's more, a fast algorithm *à la* Brandes (2001) was not easy to implement.

This concept of weak, albeit relevant bridging links, providing a network with connectivity, was instrumental to the development of the subsequent part, centered on the concept of communities of investors involved in a common, although indirect and sometimes convoluted, business. This has been long debated in the Japanese industrial economics literature, and the technical term used in the specialists' lingo is "keiretsu". Loosely speaking, it refers to a group of business partners sharing a common interest.

A brief historical digression is on order: such organizations of investors were deemed to exert their control in the style of a family-owned cartel, leading, beyond a certain degree, to quasi-monopolistic or monopolistic closed markets. Such concentration of power had been common in pre-WWII Japan; however, after 1945, the occupying forces approached this as a problem, and

attempted to disband those financial cliques.

Our a priori intuition, in the original approach, was that historical business partners were able to re-enact and disguise their pre-existing ties, even after their *ex lege* dissolution, by means of indirect and circuitous links, possibly through small intermediaries.

Intriguingly, despite years of controversy, the persistence of keiretsu organizations has been hard to prove or disprove, even by resorting to advanced econometric tools. On the other hand, through the graph-theoretical lens, a novelty in this specific field, we brought into light evidence on this phenomenon, which turned out not to be so subtle.

In this work, both unweighted and weighted networks built according to the economic ties existing between investors exhibit a well laid out structure; the visual intuition is amply confirmed by large modularity values. Looking for proof of clusterization within same-keiretsu members proved to be an almost effortless task: statistically significant evidence has been systematically found both in the expected geographical areas and manufacturing sectors.

As a third and concluding attempt, we looked again at the concept of communities, but this time with an additional dimension included: the stress, in this case, goes beyond static graphs, and is more focused on the temporal dimension. In this context, we succesfully managed to uncover a consistent description of the communities evolution through time, an intrinsically noisy process and, therefore, rather hard to deal with.

More specifically, a dynamic stream of events, the appearance and accumulation of paper citations, is examined, and presented in a readily intelligible form, given the appropriate scale.

The mesoscopic level mentioned is suggestive of the underlying idea: data are lumped together in consecutive temporal subsets (“time slices”), securing an idea of temporal continuity, yet preserving enough instantaneous information as to render local disaggregation feasible. In other words, static

snapshots of the system of interest are sampled out of the dataset; in this case, networks of journal papers linked by a co-citation measure (bibliographic coupling), which are split into homogeneous communities including similar research.

Detecting communities "locally" in the temporal sense, however, requires corrective measures in order to guarantee continuity. To make for this drawback, information from adjacent snapshots is taken into account and used to smooth out spurious transients.

Coming to the very structure of this thesis, in Chapter 1, we introduced yet another one into the already sizeable group of established centrality measures.

Despite the apparent redundancy, the added value of our contribution lies, however, in the successful effort to quantify the elusive but pivotal concept of "bridging", i.e. the effectiveness of a node in operating as a link (in the Granovetterian sense, a weak link) between homogeneous and densely connected and distinct regions of a network. In its ultimate, parsimonious incarnation, its effectiveness has been amply put to the test, both on synthetic networks expressly conceived to thwart it, and graphs obtained from real-world data; in particular, commercial flights and scientific publications.

As aforementioned, Chapter 2 consists of a novel application of existing, consolidated, yet - within the research domain in case - mostly neglected techniques, to a data panel of Japanese business entities. The main purpose, in this case, is to demonstrate how new insights, brought in from a seemingly tangential discipline, can be offered with the purpose of advancing a dispute which has been stalling over a several decades time frame.

Indirect business ties between shareholders are put to evidence by reconstructing a network of co-investments, and by detecting existing akin communities through standard modularity optimization. This provides the research with an economically sound argument to argue for the existence of

multiple centers of interest, built around investors co-owning subsidiaries.

Finally, the closing chapter, Chapter 3, approaches the problem of describing, at the appropriate mesoscopic observational level, dynamic networks, going beyond both time-agnostic community detection methods.

A set of co-citing articles within an aptly chosen research domain, *wavelets*, has been employed as the benchmark for testing the methodology proposed.

An independent, inevitably noisy, detection process is put into place for successive time spans; transient noise is detected, accounted for and cancelled by means of a backward- and forward-looking algorithm capable to spot "continuous structural streams".

The result is a clean, robust representation of the communities (in this case, research disciplines) evolving along the data time span.

As concluding remarks, what emerges from this heterogeneous body of work is that the concept itself of community is inherently prone to the application and the scope of the specific case under scrutiny.

Therefore, the quest for a "silver bullet" solution, in our opinion, might be viewed as reductive; on the other hand, the power of network analysis lies in its very flexibility, that allowed us to tailor different tools for different situations, and which has been instrumental in shaping the undisputable success of network sciences over the last decades.

Chapter 1

Detecting global bridges in networks

1.1 Introduction

Although the history of graphs as scientific objects begins with Euler's (Euler, 1736) famous walk across Königsberg bridges, the notion of 'bridge' has rarely been tackled by network theorists[1]. Among the few articles that took bridges seriously, the most famous is probably Mark Granovetter's 1973 paper on The Strength of Weak Ties. Despite the huge influence of this paper, few works have remarked that its most original insights concern precisely the notion of 'bridge' in social networks. Granovetter suggested that there might be a fundamental functional difference between strong and weak ties. While strong ties promote homogeneous and isolated communities, weak ties foster heterogeneity and crossbreeding. Or, to use the old tönnesian cliché, strong ties generate Gemeinshaft, while weak ties generates Gesellshaft (Coser, 1975). Although Granovetter does realize that bridging is the phenomenon he is looking after, two major difficulties prevented him from a direct operationalization of such concept: "We have had neither the theory nor the measurement and sampling techniques to move sociometry from

[1] We refer to the common use of the word 'bridge', and not to the technical meaning in graph theory as 'an edge whose deletion increases its number of connected components'.

the usual small-group level to that of larger structures" (*Id.*, p. 1360). Let's start from "the measurement and sampling techniques". In order to compute the bridging force of a given node or link, one needs to be able to draw a sufficiently comprehensive graph of the system under investigation. Networks constructed with traditional ego-centered and sampling techniques are too biased to compute bridging forces. Exhaustive graphs of small social groups will not work either, since such groups are, by definition, dominated by bounding relations. Since the essence of bridges is to connect individuals across distant social regions, they can only be computed in large and complete social graphs. Hopeless until a few years ago, such endeavor seems more and more reasonable as digital media spread through society. Thanks to digital traceability it is now possible to draw large and even huge social networks (Vespignani, 2009; Lazer et al., 2009; Venturini and Latour, 2010).

Let's discuss now the second point, the "theory" needed to measure the bridging force of different edges or nodes[2]. Being able to identify bounding and bridging nodes has a clear interest for any type of network. In social networks, bounding and bridging measures (or "closure" and "brokerage", as per Burt, 2005) tell us which nodes build social territories and which allow items (ideas, pieces of information, opinions, money...) to travel through them. In scientometrics' networks, these notions tell us which authors define disciplines and paradigms and which breed interdisciplinarity. In ecological networks, they identify relations, which create specific ecological communities and the ones connecting them to larger habitats.

In all these contexts, it is the very same question that we wish to ask: do nodes or edges reinforce the density of a cluster of nodes (bounding) or do they connect two separated clusters (bridging)? Formulated in this way, the bridging/bounding question seems easy to answer. After having identified

[2]In this chapter, we will focus on defining the bridgeness of nodes, but our definition can straightforwardly be extended to edges, just as the betweenness of edges is derived from that of nodes.

the clusters of a network, one should simply observe if a node connects nodes of the same cluster (bounding) or of different clusters (bridging). However, the intra-cluster/inter-cluster approach is both too dependent on the method used to detect communities and flawed by its inherent circular logic: it uses clustering to define bridging and bounding ties when it is precisely the balance of bridges and bounds that determines clusters. Remark that, far from being a mathematical subtlety, this question is a key problem in social theory. Defining internal (Gemeinschaft) and external (Gesellschaft) relations by presupposing the existence and the composition of social groups is absurd as groups are themselves defined by social relations.

In this chapter, we introduce a measure of bridgeness of nodes that is independent on the community structure and thus escapes this vicious circle, contrary to other proposals (Nepusz et al., 2008; Cheng et al., 2010). Moreover, since the computation of bridgeness is straightforwardly related to that of the usual betweenness, Brandes' algorithm (Brandes, 2001) can be used to compute it efficiently[3]. To demonstrate the power of our method and identify nodes acting as local or global bridges, we apply it on a synthetic network and two real ones: the world airport network and a scientometric network.

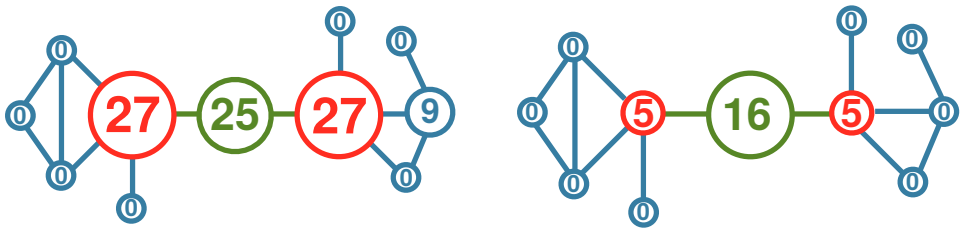
1.2 Measuring bridgeness

Identifying important nodes in a network structure is crucial for the understanding of the associated real-world system (Bonacich, 1987; Borgatti, 2005; Estrada and Rodriguez-Velazquez, 2005), for a review see Newman (2010). The most common measure of centrality of a node for network connections on a global scale is betweenness centrality (BC), which "measures the extent

[3] A plug-in for Gephi (Bastian, Heymann, and Jacomy, 2009) that computes this measure on large graphs has been developed. See Appendix A for a pseudo-algorithm useful for both node and edge bridgeness.

to which a vertex lies on paths between other vertices" (Freeman, 1977; Freeman, 1979). We show in the following that, when trying to identify specifically *global* bridges, BC has some limitations as it assigns the same importance to paths between the immediate neighbours of a node as to paths between further nodes in the network. In other words BC is built to capture the overall centrality of a node, and is not specific enough to distinguish between two types of centralities: local (center of a community) and global (bridge between communities). Instead, our measure of bridging is more specific, as it gives a higher score to global bridges. The fact that BC may attribute a higher score to local centers than to global bridges is easy to see in a simple network (Figure 1.1). The logics is that a "star" node with degree k, i.e. a node without links between all its first neighbors (clustering coefficient 0) receives automatically a $BC = k(k-1)/2$ arising from paths of length 2 connecting the node's first neighbors and crossing the central node. More generally, if there exist nodes with high degree but connected only locally (to nodes of the same community), their betweenness may be of the order of that measured for more globally connected nodes. Consistent with this observation, it is well-known that for many networks, BC is highly correlated with degree (Nakao, 1990; Goh et al., 2003; Newman, 2005). A recent scientometrics study tried to use betweenness centrality as "an indicator of the interdisciplinarity of journals" but noted that this idea only worked "in local citation environments and after normalization because otherwise the influence of degree centrality dominated the betweenness centrality measure (Leydesdorff, 2007).

To avoid this problem and specifically spot out global centers, we decompose BC into a local and a global term, the latter being called 'bridgeness' centrality. Since we want to distinguish global bridges from local ones, the simplest approach is to discard shortest paths, which either start or end at a node's first neighbors from the summation to compute BC (Eq. 1.1). This



(a) Betweenness centrality (b) Bridgeness centrality

FIGURE 1.1: The figures show the betweenness (a) and bridgeness (b) scores for a simple graph. Betweenness does not distinguish centers from bridges, as it attributes a slightly higher score (Figure a, scores = 27) to high-degree nodes, which are local centers, than to the global bridge (Figure a, score = 25). In contrast, bridgeness rightly spots out the node (Figure b, score = 16) that plays the role of a global bridge.

completely removes the paths that connect two non connected neighbors for 'star nodes' (see Figure 1.1) and greatly diminishes the effect of high degrees, while keeping those paths that connect more distant regions of the network.

More formally in a graph $\mathcal{G} = (V, E)$, where V assigns the set of nodes and E the set of links, the definition of the betweenness centrality for a node $j \in V$ stands as:

$$BC(j) = Bri(j) + local(j), \tag{1.1}$$

where

$$\begin{aligned} BC(j) &= \sum_{i \neq j \neq k} \frac{\sigma_{ik}(j)}{\sigma_{ik}} \\ Bri(j) &= \sum_{i \notin N_G(j) \wedge k \notin N_G(j)} \frac{\sigma_{ik}(j)}{\sigma_{ik}} \\ local(j) &= \sum_{i \in N_G(j) \vee k \in N_G(j)} \frac{\sigma_{ik}(j)}{\sigma_{ik}}. \end{aligned} \tag{1.2}$$

Here the summation runs over any distinct node pairs i and k; σ_{ik} represents the number of shortest paths between i and k; while $\sigma_{ik}(j)$ is the number of such shortest paths running through j. Decomposing BC into two parts

(right hand side) the first term defines actually the global term, *bridgeness centrality*, where we consider shortest paths between nodes not in the neighbourhood of j ($N_G(j)$), while the second *local term* considers the shortest paths starting or ending in the neighbourhood of j. This definition also demonstrates that the bridgeness centrality value of a node j is always smaller or equal to the corresponding BC value and they only differ by the local contribution of the first neighbours. Fig. 1.1 illustrates the ability of bridgeness to specifically highlight nodes that connect different regions of a graph. Here the BC, Fig. 1.1(a), and bridgeness centrality values, Fig. 1.1(b), calculated for nodes of the same network demonstrate that bridgeness centrality gives the highest score to the node which is central globally (green), while BC does not distinguish among local or global centers, and actually assigns the highest score to nodes with high degrees (red).

In the following, to further explore the differences between these measures we define an independent reference measure of bridgeness using a known partitioning of the network. This measure provides us an independent ranking of the bridging power of nodes, that we correlate with the corresponding rankings using the BC and bridgeness values. In addition we demonstrate via three example networks that bridgeness centrality is always more specific than BC to identify global bridges.

1.3 Computing global bridges from a community structure

To identify the global bridges independently from their score in BC or bridgeness, we use a simple indicator inspired by the well-known Rao-Stirling index (Rao, 1982; Stirling, 2007; Rafols, 2014; Jensen and Lutkouskaya, 2014), as this indicator is known to quantify the ability of nodes to connect different

communities. Moreover, it includes the notion of "distance", which is important for distinguishing local and global connections. However, we note that this index needs as input a prior categorization of the nodes into distinct communities. Our global indicator G for node i is defined as:

$$G(i) = \sum_{J \in \text{communities}} l_{IJ}\delta_{i,J} \tag{1.3}$$

where the sum runs over communities J (different from the community of node i, taken as I), $\delta_{i,J}$ being 1 if there is a link between node i and community J and 0 otherwise. Finally, l_{IJ} corresponds to the 'distance' between communities I and J, as measured by the inverse of the number of links between them: the more links connect two communities, the closer they are. Nodes that are only linked to nodes of their own community have $G = 0$, while nodes that connect two (or more) communities have a strictly positive indicator. Those nodes that bridge distant communities, for example those that are the only link between two communities, have high G values.

As a next step we use this reference measure (i.e. the global indicator) to rank nodes and compare it to the rankings obtained by the two tentative characteristics of bridging (BC and bridgeness) in three large networks.

1.4 Synthetic network: unbiased LFR

We start with a synthetic network obtained by a method similar to that of Lancichinetti, Fortunato, and Radicchi (2008). This method leads to the so-called 'LFR' networks with a clear community structure, which allows to easily identify bridges between communities. We have modified the algorithm to obtain bridges without the degree bias which arises from the original method. Indeed, LFR first creates unconnected communities and then chooses randomly internal links that are reconnected outside the community.

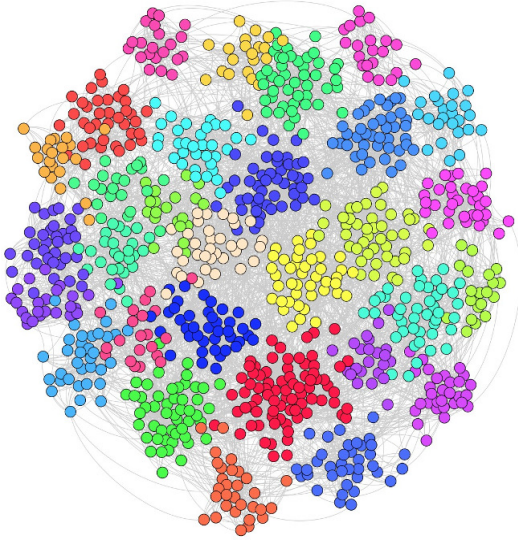


FIGURE 1.2: Artificial network with a clear community structure using Lancichinetti, Fortunato, and Radicchi (2008) method. For clarity, we show here a smaller network containing 1.000 nodes, 30 communities, 7.539 links (20% inter- and 80% intra-community links). Each color corresponds to a community as detected by modularity optimization (Newman, 2010; Blondel et al., 2008).

This leads to bridges, *i.e.* nodes connected to multiple communities, which have a degree distribution biased towards high degrees. In our method, we avoid this bias by randomly choosing nodes, and then one of their internal links, which we reconnect outside its community as in LFR. As reference, we use the global indicator defined above. As explained, this indicator depends on the community structure, which is not too problematic here since, by construction, communities are clearly defined in this synthetic network.

Fig. 1.3(a) shows that bridgeness provides a ranking that is closer to that of the global indicator than BC. Indeed, we observe that the ratio for bridgeness is higher than for BC. This means that ordering nodes by their decreasing bridgeness leads to a better ranking of the 'global' scores - as measured by G - than the corresponding ordering by their decreasing BC values. As shown in the simpler example of a 1000-node network (demonstrated in Fig. 1.2), BC fails because it ranks too high some nodes that have no external connection but have a high degree. A detailed analysis of the nodes of a

cluster is given in Supplementaty Information.

In addition we directly measured the average relative contribution of the local term (Eq. 1.4) in BC for nodes of the same degree, see Fig. 1.3(b).

$$\langle locterm \rangle_i(k) = \langle (BC(i,k) - Bri(i,k))/BC(i,k) \rangle_i \tag{1.4}$$

We observe a negative correlation, which means that the local term is dominating for low degree nodes, while high degree nodes have higher bridgeness value as they have a higher chance to connect to different communities.

1.5 Real network 1: airport's network

Proving the adequacy of bridgeness to spot out global bridges on real networks is more difficult, because generally communities are not unambiguously defined, therefore neither are global bridges. Then, it is difficult to show conclusively that bridgeness is able to specifically spot these nodes. To answer this challenge, our strategy is the following:

(i) We use flight itinerary data providing origin-destination pairs between commercial airports in the world (International Air Transport Association). The network collects 47,161 transportation connections between 7,733 airports. Each airport is assigned to its country.

(ii) We consider each country to be a distinct 'community' and compute a global indicator based on this partitioning, as it allows for an objective (and arguably relevant) partition, independent from any community detection methods. Then we show that bridgeness offers a better ranking than BC to identify airports that act as global bridges, *i.e.* that connect countries internationally.

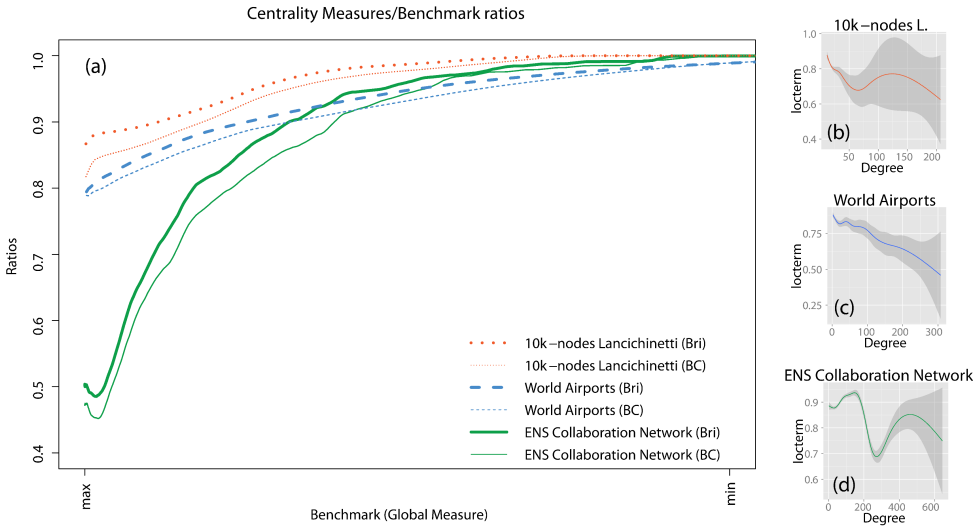
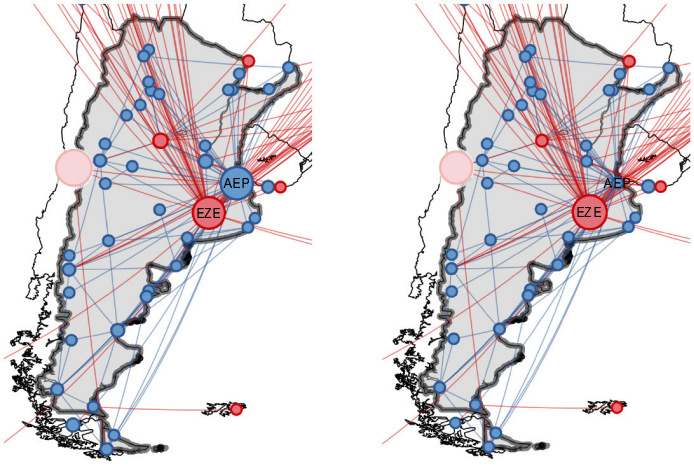


FIGURE 1.3: (a) Ability of BC or bridgeness to reproduce the ranking of bridging nodes, taking as reference the global indicator (Eq2). For each of the three networks, we first compute the cumulative sums for the global measure G, according to three sorting options: the G measure itself and the two centrality metrics, namely BC and bridgeness. By construction, sorting by G leads to the highest possible sum, since we rank the nodes starting by the highest G score and ending by the lowest. Then we test the ability of BC or bridgeness to reproduce the ranking of bridging nodes by computing the respective ratios of their cumulative sum, ranking by the respective metric (BC or Bri), to the cumulative obtained by the G ranking. A perfect match would therefore lead to a ratio equal to 1. Since we observe that the ratio for bridgeness is higher than for BC, this means that ordering nodes by their decreasing bridgeness leads to a better ranking of the 'global' scores as measured by G. To smooth the curves, we have averaged over 200 points. Curves corresponding to different networks are colorised as LFR (red), Airports (blue), ENS (green). (b, c, d): average relative local terms as function of node degree for the three investigated networks (for definition see text).



(a) Betweenness centrality (b) Bridgeness centrality

FIGURE 1.4: Example of the two largest Argentinean airports, Ezeiza (EZE) and Aeroparque (AEP). Both have a similar degree (54 and 45 respectively), but while the first connects Argentina to the rest of the world (85% of international connections, average distance 2,848 miles, G=2,327.2), Aeroparque is only a local center (18% of international connections, average distance 570 miles, G=9.0). However, as in the simple graph (Figure 1), BC gives the same score to both (BC_{EZE}=79,000 and BC_{AEP} = 82,000), while bridgeness clearly distinguishes the local center and the bridge to the rest of the world, by attributing to the global bridge a score 250 times higher (Bri_{EZE}=46,000 and Bri_{AEP} = 174). Red nodes represent international airports while blue nodes are domestic.

As an example, in Fig. 1.4 we show the two largest airports of Argentina, Ezeiza (EZE) and Aeroparque (AEP). Both have a similar degree (54 and 45 respectively), but while the first connects Argentina to the rest of the world, Aeroparque mostly handles domestic flights, thus functioning as a local center. This is confirmed by the respective G values: 2,327.2 (EZE) and 9.0 (AEP). However, just like in our simple example in Fig. 1.1, BC gives the same score to both, while bridgeness clearly distinguishes between the local domestic center and the global international bridge by attributing to the global bridge a score 250 times higher (see Fig. 1.4). This can partly be explained by the fact that AEP is a 'star' node (low clustering coefficient: 0.072), connected to 12 very small airports, for which it is the only link to the whole network. All the paths starting from those small airports are cancelled in the computation of the bridgeness (they belong to the *local* term in Eq. 1.1), while BC counts them equally as any other path.

More generally, Fig. 1.3 shows that, as for the Airport network, bridgeness provides again a ranking that is closer to that of the global indicator. Indeed, ordering nodes by their decreasing bridgeness leads to a ranking that is closer to the ranking obtained by the global score than the ranking by decreasing BC. In addition, we found again negative correlations between the average relative local term and node degrees, see Fig. 1.3(c), assigning similar roles for low and high degree nodes as in case of the synthetic network.

1.6 Real network 2: scientometric network of ENS Lyon

The second example of a real network is a scientometric graph of a scientific institution (Grauwin and Jensen, 2011), the "Ecole normale supérieure

de Lyon" (ENS, see Figure 1.5). This networks adds authors to the usual co-citation network, as we want to understand which authors connect different sub-fields and act as global, interdisciplinary bridges. To identify the different communities, we rely on modularity optimization (Blondel et al., 2008), which leads to a relevant community partition because scientific networks are highly structured by disciplinary boundaries. This is confirmed by the high value of modularity generated by this partition (0.89). In Figure 1.5, the authors of different communities are shown with different colors, and their size corresponds to their betweenness (left) or bridgeness (right) centrality, which clearly leads to highlight different authors as the main global bridges, which connect different subfields. We compute the Stirling indicator (Eq. 1.1) based on the modularity structure to identify the global bridges. As for the previous networks, Fig. 3 shows that bridgeness ranks the nodes in a closer way than BC to the ranking provided by the global measure based on community partition. On the other hand the corresponding $\langle locterm \rangle(k)$ function, see Fig. 1.3(d), suggests a slightly different picture in this case. Here nodes with large but moderate degrees (smaller than ~200) have high local terms suggesting that they act as local centres, while nodes with higher degrees have somewhat smaller local terms assigning their role to act as global bridges.

1.7 Discussion

In this chapter, we introduced a measure to identify nodes acting as global bridges in complex network structures. Our proposed methodology is based on the decomposition of BC into a local and global term, where the local term considers shortest paths that start or end at one of the node's neighbors, while the global term, what we call bridgeness, is more specific to identify nodes which are globally central. We have shown, on both synthetic and real

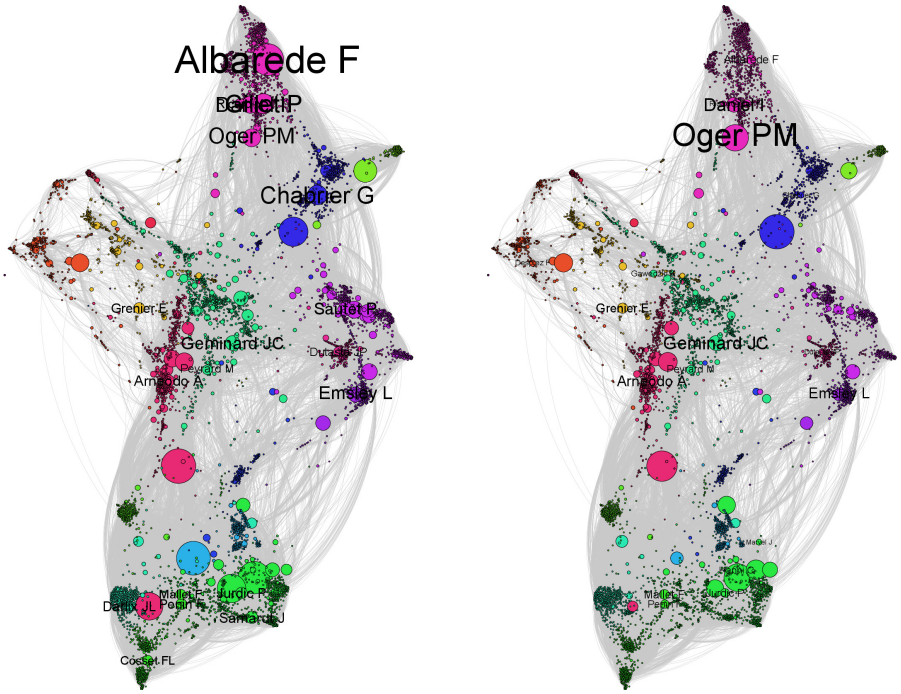


FIGURE 1.5: Co-citation and co-author network of articles published by scientists at ENS de Lyon. Nodes represent the authors of references appearing in the articles, while links represent co-appearances of these features in the same article. The color of the nodes corresponds to the modularity partition and their size is proportional to their BC (left) or to their bridgeness (right), which clearly leads to different rankings (references cited are used in the computations of the centrality measures but appear as dots to simplify the picture). We only keep nodes - authors - that appear on at least four articles and links that correspond to at least 2 co-appearances in the same paper. After applying these thresholds, the 8,000 articles lead to 8,883 nodes (author or references cited in the articles) and 347,644 links. The average degree is 78, the density 0.009 and the average clustering coefficient is 0.633. Special care was paid to avoid artifacts due to homonyms. Weights are attributed to the links depending on the frequency of co-appearances (as cosine distance, see Grauwin and Jensen, 2011).

networks, that the proposed bridgeness measure improves the capacity to specifically find out global bridges as it is able to distinguish them from local centers. One crucial advantage of our measure of bridgeness over former propositions is that it is independent of the definition of communities.

However, the advantage in using bridgeness depends the precise topology of the network, and mainly on the degree distribution of bridges as compared to that of all the nodes in the network. When bridges are high-degree nodes, BC and bridgeness give an equally good approximation, since high-degree bias do not play an important role in this case. Instead, when some bridges have low degrees, while some high-degree nodes act like local centers of their own community, bridgeness is more effective to identify bridges as BC gives equally high rank to nodes with high degree, even if they are not connected to nodes outside of their community. We demonstrated that bridgeness is systematically more specific to spot out global bridges in all the networks we have studied here. Although the improvement was small on average, typically 5 to 10%, even a small amelioration of a widely used measure is in itself an interesting result.

We should also note that, except on simple graphs, comparing these two measures is difficult since there is no clear way to identify, independently, the 'real' global bridges. We have used community structure when communities seem clear-cut, but then we fall into the circularity problems stressed in the introduction. Using metadata on the nodes (*i.e.* countries for the airports) may solve this problem but raises others, as metadata do not necessarily correspond to structures obtained from the topology of the network, as shown recently on a variety of networks (Hric, Darst, and Fortunato, 2014). Another possible extension would be to identify overlapping communities to identify independently global bridges, as nodes involved in multiple communities, and correlate them with the actual measure, which provides a direction for future studies. However, in any case identifying global bridges remains a

difficult problem as it is tightly linked to another difficult problem, that of community detection. Decomposing BC into a local and a global term helps to improve the solution, but many questions remain still open for further inquiry.

Chapter 2

The evolution of Japanese business networks in ASEAN countries since the 1960s

2.1 Introduction

Network analysis techniques remain rarely used for understanding international management strategies. This chapter highlights their value as research tool in this field of social science using a large set of micro-data (> 20.000) to investigate the presence of networks of subsidiaries overseas. The research question is the following: to what extent did/do global Japanese business networks mirror organizational models existing in Japan?

In particular, we would like to assess how much the links building such business networks are shaped by the structure of big-size industrial conglomerates of firms headquartered in Japan described, sometimes imprecisely, as horizontal keirestsu (HK).

As per Lawrence's (1993) factual definition, these are groups which "typically include a lead bank, a trust bank, life and casualty insurance companies, and a general trading company". He adds, "members of the group exchange shares, have interlocking directorates, jointly appoint officers and other key

personnel, hold regular meetings of company presidents and engage in joint investment undertakings in new industries". Finally, concerning credit access, characteristically "lead banks may provide group members with loans at preferential rates and with implicit assurances to stand behind them in times of trouble".

The major part of the academic community in the fields of management and industrial organization considers that formal links can be identified among firms belonging to HK. Miwa and Ramseyer (Miwa and Ramseyer, 2002; Miwa and Ramseyer, 2006) challenge this claim and argue that the evidence supporting the existence of HK is weak.

So far, quantitative empirical investigation has been conducted exclusively using data for firms incorporated in Japan. Our study tests the Miwa-Ramseyer hypothesis (MRH) at the global level using information on the network of Japanese subsidiaries overseas. We identify linkages among Japanese subsidiaries overseas using an objective criterion: the subsidiaries share at least two Japanese co-investors (firms headquartered in Japan). The results obtained lead us to reject the MRH for the global dataset, as well as for subsets restricted to the two main regions/countries of destination of Japanese foreign investment: China (broadly defined as to include Hong Kong and Taiwan), and Southeast Asia. The results are robust to the weighting of the links, with different specifications, and are observed in most industrial sectors; the main exception is the automotive industry for which a straightforward explanation (unrelated to the MRH) exists. The global Japanese network became increasingly complex during the late 20th century as a consequence of increase in the number of Japanese subsidiaries overseas but the key features of the structure remained rather stable. We draw implications of these findings for academic research in international business and for professionals involved in corporate strategy.

The heuristic and explanatory power of network analysis techniques is

widely acknowledged in various disciplines of social science (Scott, 1999; Padgett and Powell, 2012). They remain however rarely used for the empirical analysis of international management strategies, except in a few recent studies based on relatively small samples (e.g. Shi et al., 2014; for a review, see Hoang and Yi, 2015). This remark applies also to academic research in management but also, surprisingly, to professional strategic consulting and business auditing activities. The results presented in the volume edited by David and Westerhuis (2014) demonstrate that a strong interest is emerging for studies with country-level historical perspective, including on Japan (Koibuchi and Okazaki, 2014).

This chapter brings into such context network analysis tools with a study using a large sample of micro-data for the global network of Japanese subsidiaries overseas and, through our analysis, we practically demonstrate the interest and the still untapped potential of network techniques in international business.

The topic we investigate is related to an unsettled issue in Japanese business history that remains entirely relevant for analyzing present-day Japanese business strategies at home and abroad.

Specifically, we investigate if the structure of Japanese business networks is reminiscent of the corresponding horizontal conglomerates (HK). The majority of the academic community in the fields of management and industrial organization considers that the links between firms belonging to these HK can be identified through information on main-bank, cross-ownership, and transactions (e.g. Gerlach, 1992; Aoki and Saxonhouse, 2000). Notably, the very existence of HK is cast into question by Miwa and Ramseyer (Miwa and Ramseyer, 2002; Miwa and Ramseyer, 2006) who criticize the weakness of the evidence proving it.

Alternatively, they interpret it as an ideological construct that was firstly

devised by Japanese Marxists in the 1950s to be, later on, adopted by the Dodwell, a marketing company, and which was finally endorsed by non-Marxist scholars as well. As a concluding remark, it is worth stressing that quantitative empirical investigation have been conducted, up to now, exclusively using data for firms incorporated in Japan.

In this chapter the Miwa-Ramseyer hypothesis (MRH) is tested globally, using information on Japanese subsidiaries overseas. The data are obtained from a nearly exhaustive global survey of Japanese overseas subsidiaries conducted by a private company, the Toyo Keizai Shinposha. Their dataset (henceforth TKZ) includes more than 20,000 firms in total, of which around 6,000 in the manufacturing sector, the one relevant to our analysis. The coverage is global and includes all recipient countries of Japanese foreign direct investment. The TKZ database reports information for wholly owned companies or joint ventures with local partners. Available information enables identifying Japanese and non-Japanese investors, and the shares owned by each firm. Membership of Japanese parent companies in one of the HK is defined on the basis of two indicators also supplied by TKZ: involvement regular meetings and equity ownership by firms identified core members of the HK. We use community detection techniques with different specifications and subsets of data in order to assess robustness of our results. Although with some caveats, the results obtained lead us to reject the MRH.

The remainder of the chapter is organized as follows: section 2.2 offers an overview of the state of the art in business network analysis, and identifies some major gaps in the literature; section 2.3 describes the TKZ dataset; section 2.4 describes the hypotheses under investigation; section 2.5 describes the strategy adopted to test the MRH and discusses the results; section 2.6 summarizes the findings and elaborates on their implications for scholars and managers.

2.2 Non-Japanese and Japanese business networks: state of the art and gaps in the literature

The major part of the studies on business networks remain focused on the analysis of relatively simple networks, either of centered on one particular firm, or using a small sample of observations.

What are the reasons for this limited development of complex network analysis in management, compared to other fields of social science, in particular economics or sociology? The three major intertwined explanations seems related to the strong focus typical of studies in management: (a) qualitative techniques (Harvard Business School type case studies), (b) intra-firm networks, and (c) directed networks (as opposed to undirected networks that are usually the more complex ones).

First, qualitative techniques enable analyzing extremely complex conditions, structures and strategies, but usually for one particular firm or, at best (from the viewpoint of network analysis), for a relatively small group of firms (e.g. Forsgren and Johanson, 1992). In some minority cases, the scope is a little wider since the interactions of financial holdings and other investors are taken into account on the basis of information of equity holding.

The focus on intra-firm networks is a legacy of a tradition in management going back to Chandler (1962, 1977) that analyses the firm structural changes in response to the top management decisions meant to adjust corporate business strategies.

Finally, the focus on directed networks is also perfectly understandable: business networks are viewed essentially as centered on a group of top managers. The types of inter-firms networks considered are mostly hierarchical pyramids of firms linked by equity ownership relations with different tiers corresponding to subsidiaries of first, second, third, or Nth rank. An interest for studies taking into account both strong and weak business ties using

network analysis is however emerging recently (Kilkenny and Fuller-Love, 2014).

In this context, the specialized sub-field of studies on Japanese business networks stands aside. The theory of the Japanese firm as a nexus of treaties formulated by Aoki (1984a, 1984b) emerged at the time when Japanese business networks were barely discussed (the book by Kono, 1984, on the strategy and structure of Japanese enterprises, one of the most widely circulated in English around that period, does not mention business networks at all). Aoki's game-theoretical approach, although concentrated on intra-firm aspects, was very influential, as it provided an analytical framework applicable to undirected inter-firm networks. This solid conceptual bases enabled the development of an empirical stream of research that was later identified as keiretsu studies (in particular Gerlach, 1992; Lincoln, Gerlach, and Takahashi, 1992; Lincoln, Gerlach, and Ahmadjian, 1996; Weinstein and Yafeh, 1995; Aoki and Saxonhouse, 2000; Nakamura, 2002; Lincoln and Gerlach, 2004. McGuire and Dow (2009) provides the most extensive recent survey of this stream of research.

The Japanese term keiretsu, which is usually translated as 'alignment', indicates that a firm has a set of preferential cooperations with another firm, generally bigger and in that case the relation is clearly hierarchical, or with a group of firms. The relation of the firm with this group can be either hierarchical, as aforementioned, or not hierarchical, in this latter case the firm is a member of an undirected network. In the case of hierarchical relations between firms, the structure is the same as in vertically organised business groups that exist in all regions of the world. This pattern is described in Japanese business studies as vertical keiretsu (VK), while the undirected network is described as horizontal keiretsu (HK). It should be noted, however, that there are overlaps between VK and HK. Specialized manufacturing groups sur as Toyota and Mitsubishi are vertically organized and, at the same

time, they are part of an HK (Mitsui and Mitsubishi keiretsu, respectively.)

The development of the keiretsu studies led, among other consequences, to revive the interest in Japanese business history studies investigating the strategies and structures of prewar groups owned by kinship networks, i.e. zaibatsu such as Mitsui, Mitsubishi, Sumitomo, and Yasuda, dissolved in 1946 upon request of the U.S. occupation authorities. An obvious issue was assessing the strength of post- war links between companies that belonged to these pre-war groups and that morphed in the 1950s into HK type conglomerates whose membership was somehow different from the Zaibatsu's (see also Ramseyer and Miwa, 2007) for a discussion on the dissolution of pre-war zaibatsu). Ironically, the expansion of the literature on HK accelerated in the 1990s and 2000s, precisely at the time when the ties were becoming increasingly informal and weak. What is more, the description of the structure of HK became obsolete as a consequence of waves of mega-mergers of the major Japanese banks that took place in 2006. This led some of the key figures in the keiretsu studies to reflect on their demise ("why they are gone?") and the future of Japanese business groups (e.g. Lincoln and Shimotani, 2009; see also McGuire and Dow, 2009 for a similar discussion).

Looking at studies concerning Japanese business network overseas, we can observe that a number of papers are explicitly referring to keiretsu membership for investigating various issues such investment overseas (Belderbos and Sleuwaegen, 1996), as spatial location decision and agglomeration effects (e.g. Belderbos and Carree, 2002; Yamashita, Matsuura, and Nakajima, 2014; Zhang, 2015). A few papers also discuss the importance of keiretsu in shaping Japanese business networks overseas, in some case using the large datasets such as the TKZ database (e.g. Zhang, 2015). However, to the best of our knowledge, there has been no attempt so far to use standard network analysis techniques to unravel the structure of Japanese networks overseas and our study aims at filling this gap.

2.3 Dataset and descriptive statistics

The TKZ dataset we use actually is composed by two datasets, both produced by the Toyo Keizai Shinposha, a private company whose denomination in English is Oriental Economist (http://corp.toyokeizai.net/en/). One of the attractive features of this database is that it has not been constructed by the Japanese government or a not-for-profit semi-public body, but rather by a private company: henceforth, their quality and accuracy was meant to generate a positive return on investment. Thus, that these volumes are rather best sellers than confidential publications is indicative of such quality and of the trust the public had in the information supplied. The surveys have been repeated yearly since more than 50 years on the basis of voluntary participation.

The first database reports micro-data resulting from a yearly survey administered in 2005 to Japanese subsidiaries based overseas (a sizeable 20,700), circulated in the 2006 TKZ edition that we use for reasons explained below. The firms respond on a voluntary basis, and some piece of information is sporadically missing in the returned forms, collected and processed by Toyo Keizai. The sampling rate is not disclosed by the Toyo Keizai, but the consensus is that coverage is extremely high because the respondents are not expected to report confidential information. The list of data requested is limited to the denomination, address, industrial sector, paid-up capital, name and share of each Japanese equity owner, and share of local investors, when joint ventures are established with local foreign partners; notably, the respondents are not required to disclose the identity of the local partners. Toyo Keizai processes the information as to include a unique code for each subsidiary and, more importantly, for equity owners (companies headquartered in Japan; no code for local investor). The equity owners univocal coding system aptly implemented by Toyo Keizai prevents any risk of confusion due

to mislabelling: the local subsidiaries managers responding to the survey are held back from using possibly inconsistent textual denominations. Since we had the chance to access the electronic version of the 2006 database (more often than not, the printed version is used), there is no risk of error or omission (at least not by our research team).

We process the information available in the database to identify conglomerates of Japanese overseas businesses, by defining a quantitative and objective criterion: two Japanese investors (firms headquartered in Japan) are considered to be linked (as nodes connected by an edge on the graph) if they co-invest in, i.e. they co-own, one or more overseas subsidiaries.

These co-ownership relations are of utmost relevance since they are measurable in stark contrast with relations between firms only of informal cooperation and/or repeated transactions, without any equity ownership tie.

As side note, a subsidiary may be involved in more than one business network. Thus, such shared subsidiaries heuristically play the role of "bridges" between two networks, therefore they contribute to the network cohesion. For the interested reader, such links are related to the concept of weak ties (Granovetter, 1973).

Furthermore, the advantage, among others, of the TKZ dataset is to include information on mid-size groups. However, after 2006, the mega-mergers of some Japanese banks disrupted such fine scale structure and for this reason we focus our analysis on the pre-2006 networks. The implications of our results on pre-2006 networks for analyzing present day conditions are discussed in section 2.6.

The second Toyo Keizai database we use, much smaller, has been obtained from the last issue (published in 2000) of the Toyo Keizai "Keiretsu Survey" (a distinct product from the Toyo Keizai yearbook on industrial groups, published yearly). It provides very valuable information on the 6 big

HK: Mitsui, Mitsubishi, Sumitomo, Sanwa, Fuyo, and Ikkan, that can be considered as the indirect heirs of kinship-based zaibatsu ("financial cliques") such as Mitsubishi, Mitsui, and Sumitomo, that were dissolved in 1947. Relying on information obtained from Toyo Keizai Shinposha (2000), we retain two criteria for membership in one of the 6 HK: (i.) the firm is a member of one of the 6 Chief Executive Officers "clubs" (one per HK) meeting on a weekly/monthly basis; (ii.) the firm is among the top 50 companies by share of equity ownership of companies members of one of the 6 "clubs". To the best of our knowledge, the Toyo Keizai did not officially explain why the "Keiretsu Survey" has been discontinued, the 2000 issue being the last one. Two alternative interpretations can be considered. The first one, which can be considered consistent with the Miwa-Ramseyer view, is that the Toyo Keizai finally acknowledged the fallacy in the keiretsu existence nowadays. The second one, which we tend to favor, is that drastic changes in the organizational structure of Japanese business networks (in Japan, and - presumably and consequently - abroad) resulted in a dying out of the preferential links and cooperation networks that had been identified earlier. The information reported in the keiretsu volume was therefore becoming less relevant and, at any rate, redundant with the one provided in the separate Toyo Keizai volume on business groups, more focused on the concept of vertical keiretsu. In particular, as aforementioned, the 2006 mega mergers of a number of Japanese banks rendered less and less relevant keeping a list of firm memberships in a "club".

Nevertheless, the 2000 database provided by the latest issue of the "Keiretsu Survey" is valuable to our study as it provided us with the backbone underlying the keiretsu web we test our hypothesis against.

Thus, summing up, the two dataset together concur to build a coherent perspective: from the first dataset we extract with network analysis tools, described in Sec. 2.5, the granularity of the Japanese business network, i.e.,

its communities structure; on the other hand, we compare this empirical evidence of communities against the "Keiretsu Survey" dataset to infer if such communities mirror the HK organization.

2.3.1 Worldwide waves of investment

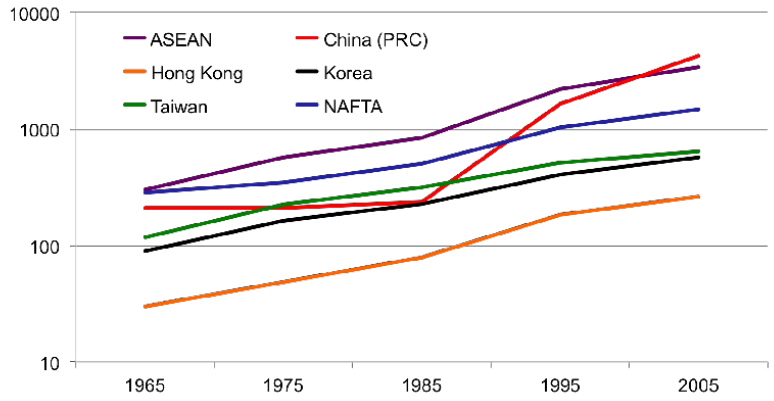
Japanese foreign direct investment (FDI) did not took place in parallel in all regions of the world, but rather in successive waves. Japanese investors were first attracted in the 1960s and 1970s by the comparative advantages of ASEAN countries, Hong Kong and Taiwan, and the possibility to gain access to these emerging markets. From the 1970s, and especially during the 1980s, North America and Europe also became important destinations. Finally, the gradual opening of the People's Republic of China to international trade and foreign investors in the 1980s resulted in a reorientation of Japanese foreign direct investment that accelerated in the 2000s; China became the main target country in terms of flows, and - after some time lag - also in terms of stock. Evidence from the TKZ is presented in Figure 2.1.

2.3.2 Heterogeneity in ownership and prevalence of manufacturing

The ratio of TKZ subsidiaries which are included and used in reconstructing the Japanese investors business networks vary manifestly by country. This is both because of the overall quota of manufacturing enterprises (very low in the EU, high in ASEAN countries, cfr. Figure 2.2), and the ratio of single owned subsidiaries (very high in the EU, lower in ASEAN countries, cfr. Tables 2.4, 2.5, below).

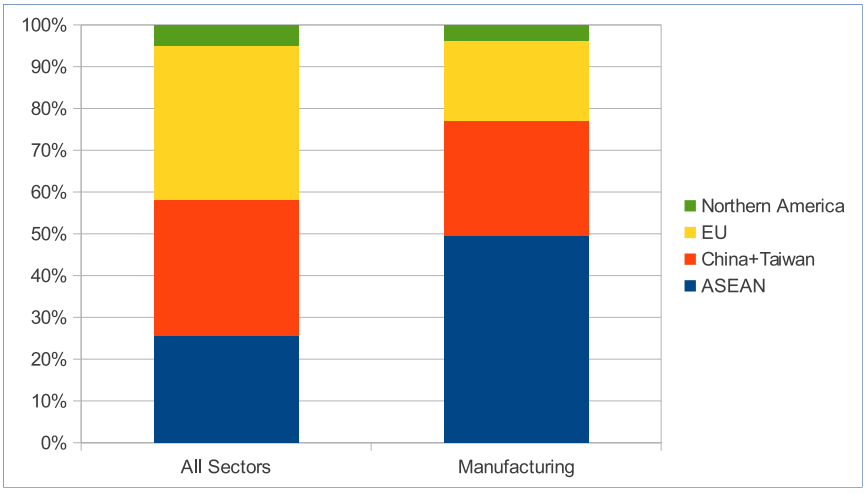
As mentioned previously, non-manufacturing businesses are not included. This is ostensibly visible in the most extreme cases, the Netherlands and Indonesia respectively (cfr. Table 2.1).

FIGURE 2.1: Japanese co-investment overseas (number of firms in the Toyo Keizai database for 1965, 1975, 1985, 1995, and 2005; log scale)



Note: The 10 countries of the ASEAN (Association of Southeast Asian Nations) are: Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam; NAFTA: North America Free Trade Agreement area (Canada, Mexico, United States).

FIGURE 2.2: % of Japanese FDI in USD (source: TKZ), by macroarea, all sectors vs. manufacturing only



Country	TOT	INCL	DIFF%
China	4.414	3.497	-21%
USA	3.419	1.194	-65%
Thailand	1.534	1.116	-27%
Hong Kong	1.109	254	-77%
Singapore	1.036	276	-73%
Taiwan	907	520	-43%
United Kingdom	806	188	-77%
Malaysia	774	516	-33%
Korea	682	453	-34%
Indonesia	673	571	-15%
Germany	637	126	-80%
Philippines	453	274	-40%
Australia	402	69	-83%
France	380	123	-68%
Netherlands	358	56	-84%
Canada	266	91	-66%
Brazil	261	137	-48%
Viet Nam	250	200	-20%
Mexico	219	124	-43%
India	199	157	-21%
(61 more countries omitted)	...	...	...
TOT	20.488	10.481	-49%

TABLE 2.1: Frequencies of firms by country: full TKZ dataset (TOT), firms included (INCL), % difference; first 20 countries sorted by decreasing TOT

Area	NUMEMPL	K
ASEAN	518	21.980
China and Taiwan	366	12.204
EU	331	20.191
Northern America	401	37.900

TABLE 2.2: Average number of employees (NEMPL) and capitalization in USD (K), disaggregated by macroarea

Industrial sector	Code	NUMEMPL	K
Food	600	346	24.781
Textiles	700	380	5.117
Wood and Furniture	800	323	12.252
Pulp and Paper	900	246	31.144
Publishing and Printing	1000	159	28.300
Chemicals and Pharma	1100	167	18.894
Petroleum and Coal (incl. Plastics)	1200	55	66.326
Rubber and Leather	1300	258	6.185
Refractories and Glass	1400	477	6.859
Steel	1500	313	140.409
Non-ferrous Metals	1600	545	37.619
Metal Products	1700	218	11.905
Machinery	1800	236	11.244
Electric/Electronic Devices	1900	832	27.436
Transp. Mach. and Shipbuilding	2000	308	28.089
Automobiles and Parts	2100	563	31.848
Precision Machinery	2200	541	11.677
Miscellaneous Manufacturing	2300	253	4.531

TABLE 2.3: Average number of employees (NEMPL) and capitalization in USD (K), disaggregated by industrial sector

2.3.3 Heterogeneity in subsidiaries size and capitalization

As it will become noticeable in the analysis, geographical and sectoral peculiarities are evident in the dataset, and play a relevant role in shaping the topological features of networks built from disaggregated data. In Tables 2.2, 2.3, descriptive data are shown over both dimensions.

Area	Avg	1	2	3	4	5	(%) 2+
ASEAN	1.66	1530	553	259	99	58	39%
China and Taiwan	1.52	2190	774	281	86	37	35%
EU	1.21	648	102	20	6	1	17%
Northern America	1.28	967	189	55	8	3	21%

TABLE 2.4: Co-investors per subsidiary, disaggregated by geographical area

Industrial sector	Code	Avg	1	2	3	4	5	(%) 2+
Food	600	1.59	306	101	48	18	7	36%
Textiles	700	1.83	252	151	74	30	7	51%
Wood and Furniture	800	1.57	52	14	5	1	4	32%
Pulp and Paper	900	1.81	47	34	15	1	3	53%
Publishing and Printing	1000	1.11	50	4	1	0	0	9%
Chemicals and Pharma	1100	1.52	995	343	120	40	18	34%
Petroleum and Coal (incl. Plastics)	1200	2.05	9	3	4	3	0	53%
Rubber and Leather	1300	1.41	177	42	24	4	0	28%
Refractories and Glass	1400	1.42	174	53	17	5	1	30%
Steel	1500	2.27	74	50	39	15	16	62%
Non-ferrous Metals	1600	1.59	155	43	21	10	5	34%
Metal Products	1700	1.63	261	112	40	22	5	41%
Machinery	1800	1.38	803	194	55	16	13	26%
Electric/Electronic Devices	1900	1.26	1518	272	68	14	8	19%
Transp. Mach. and Shipbuilding	2000	1.54	43	20	8	1	0	40%
Automobiles and Parts	2100	1.64	767	339	135	42	20	41%
Precision Machinery	2200	1.15	224	30	2	0	1	13%
Miscellaneous Manufacturing	2300	1.23	279	41	10	3	1	16%

TABLE 2.5: Co-investors per subsidiary, disaggregated by industrial sector

2.3.4 Co-investments

Globally, every subsidiary is owned by approximately one and a half (1.48) investors, unevenly split between Japanese (1.31) and local (0.17). In Table 2.4, this information is disaggregated by geographical macro-area; the European Union, where less than one fifth (17%) of the subsidiaries are owned by two or more investors, appears to be the area with the least average number of investors, closely followed by Northern America.

2.4 Hypotheses

Before starting our analysis, we would like to briefly sketch the underlying hypotheses to our approach that guided us and which were, as we explain in the following, mostly driven by the available information provided by the dataset, as in the case of the geographical distribution of the Japanese foreign investments, and by sensible insightful observations on the nature of the dataset.

Hypothesis 1: considering the manufacturing firms included in the TKZ dataset in all countries for 2005, we reject the strong form of the Miwa-Ramseyer hypothesis.

With regard to the Miwa-Ramseyer Hypothesis (MRH), we adopt an agnostic view. Indeed, we accept their claim that the empirical evidence supporting the existence of HK is rather weak. These business groups would be particularly difficult to identify should the affiliations be informal, implying that they would not require any kind of binding and irreversible commitment. Moreover, anecdotal evidence indicates that a number of firms that were identified as informal members of one HK gradually shifted to an equally informal affiliation with another HK. What is more, a number of firms loosely related to a HK eventually moved to a position of dual affiliation.

Therefore, it is not surprising that the evidence obtained using panel data analysis or similar econometric techniques could be disappointing. However, we do not reject the possibility that a nexus of bilateral or multilateral treaties and repeated transactions between firms, as well as information exchanges, involvement in joint R&D projects, and cooperation in joint ventures at home and abroad could result in the formation of an indirect business network involving tightly knit clusters of firms. Furthermore, one of the main advantages of network analysis in this context is to be consistent

with the possibility of informal affiliation and/or changes in affiliation. Our first hypothesis is that Japanese business networks overseas tend to replicate familiar structures already in place in Japan when HK had very few foreign subsidiaries, that is in the 1960s and 1970s. Accordingly, what we would describe as a strong form of MRH ("[...] at root, the keiretsu do not exist." in Miwa and Ramseyer prose), can be rejected.

Hypothesis 2: the structure of Japanese business networks overseas is becoming gradually more complex but the identified key-players remained essentially the same ones during the period 1975-2005.

To what extent did the structure of Japanese business networks overseas evolved over time? The firms setting up foreign subsidiaries in the 1960s and 1970s have been overwhelmingly the biggest players in their industry, and in their respective HK, if we believe the proponents of the strong form of the horizontal keiretsu hypothesis (HKH). It is only with the dramatic increase in volume of foreign direct investment that mid-size firms (or Japanese-based firms situated in the periphery of the HK, according to the the strong form of HKH proponents) became present as parent companies of foreign subsidiaries. Since our measure of involvement in a Japanese business network overseas is defined precisely as the co-investment in foreign subsidiaries with other Japanese investors, we expect to find a high level of stability in the structure observed using information successive benchmark years. We selected four benchmark years with a 10-year interval between them: 1975, 1985, 1995, and 2005.

Hypothesis 3: consistent results are expected with or without weighting.

Weighted networks may provide an alternative insight as to the business links used to identify groups of firms on the network. Instead of mere binary yes/no relationships, a metric can be put in place to define the links between any two investors, in terms of a continuum measuring the strength of their economic ties. For instance, when considering the shared set of subsidiaries

common to two investors, as an example of the weight, both the capitalization and the owned share must be taken into consideration. See section 2.5.3 for details.

Hypothesis 4: Similar results (rejection of the MRH) are expected, even when considered separately, for three out of the four main (in terms of destination of Japanese foreign investment) world macro-regions: ASEAN, China (including Hong Kong and Taiwan), and North America. Not enough observations are available for Europe, on the other hand, in order to present conclusive results.

As observed in Section 2.3.1, foreign investments for Japanese firms historically proceeded by successive waves, first hitting ASEAN countries, to then spread to North America and Europe and, finally, to China. However, if the architecture of the Japanese business networks in these different regions of the world was determined by characteristics of the links between Japanese parent companies, we would expect to find similar results.

Hypothesis 5: When considering the peculiar structure of some industrial sectors, similar results may not be observed in all of them.

The investors and subsidiaries involved in the Japanese business networks overseas specialised in different lines of business. That is also the case of the big conglomerates (HK). However, the major part of the co-investments are likely to associate firms of the same industrial sectors, in particular in the manufacturing sector. Depending on the number of potential partners and the advantages derived from co-investments in foreign subsidiaries in terms of information sharing and risk mitigation, it is conceivable that firms that would be normally competitors may decide to cooperate in order to penetrate a foreign market. It is therefore plausible that such a specific pattern, contradicting the general trend, is observed in some sectors, e.g. automotive, food and textile.

2.5 Methodology

2.5.1 Community detection on inferred business network

We test the MRH using a standard network community detection technique (Blondel et al., 2008) on the reconstructed co-investment graph described in Sec. 2.3, matching the communities each node (firm) is assigned to, and information on real-world business conglomerates (keiretsu), in order to evaluate to what extent the communities detected in the networks correspond to communities defined by at least one of the criteria of membership of one of the 6 big HK obtained from Toyo Keizai Shinposha, 2000.

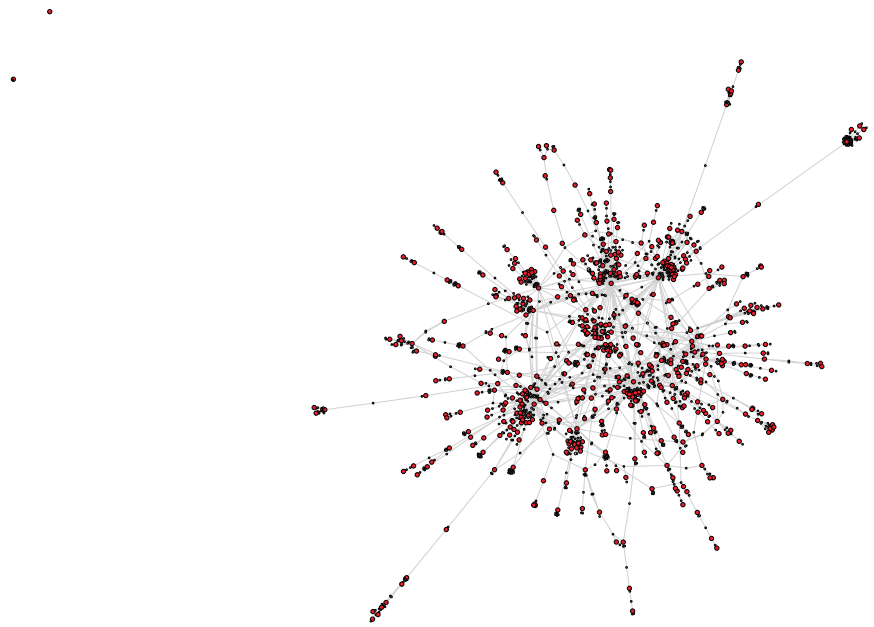
The results are tested against networks which are equivalent, degree-wise, but with randomized structure (Configuration Models, see Newman, 2003, as null hypotheses). Absence of significant correlation between communities and keiretsu is consistently shown in the latter case.

The network is built using Japanese investors appear as nodes, which are linked if they share investments in at least N subsidiaries. Figure 2.3 offers an insight of the role of subsidiaries in structuring clusters of Japanese investors.

Both parent companies in Japan and subsidiaries overseas are reliably identified by unambiguous Toyo Keizai codes. Data on foreign local partners is also available, but their identification is hardly consistent due to the absence of unique identifiers; many generic "anonymous" nodes also appear. At any rate, their contribution to the network structure appears negligible (cfr. Sec. 2.3.4), which led us to finally omit this information in order to avoid unpredictable bias.

The main purpose of representing TKZ microdata as a network of co-investments is to look for "more densely connected" sets of firms which are, by definition, communities. Firms tend to cluster together in the network structure if there are prevalent and privileged economic links between them. If we observe a non-random distribution (systematic overrepresentation) of

FIGURE 2.3: Heterogeneous network, including Japanese investors and overseas subsidiaries, ASEAN countries only



Note: red and grey nodes represent, respectively, Japanese investors and subsidiaries; sizes are arbitrary and equal within each category, in order to better reveal the underlying network structure. Zoomable version available at http://perso.ens-lyon.fr/matteo.morini/jpbusnet/Het_ic.pdf ; industrial codes (see Table 2.5) are readable as labels in the online version.

firms known to belong to a keiretsu across communities, we can conclude that the economic structure revealed by the network topology is driven by keiretsu-type inter-firm bonds.

Our hypothesis can be tested by contrasting two different characteristics of every firm: on the one hand, a firm can be a member of either one of the Big 6 keiretsu groups; on the other, it belongs to one of the communities derived from the network structure. We seek to verify the independence (or lack thereof) between the economic network, as resulting from the community detection method described above, and the keiretsu structure.

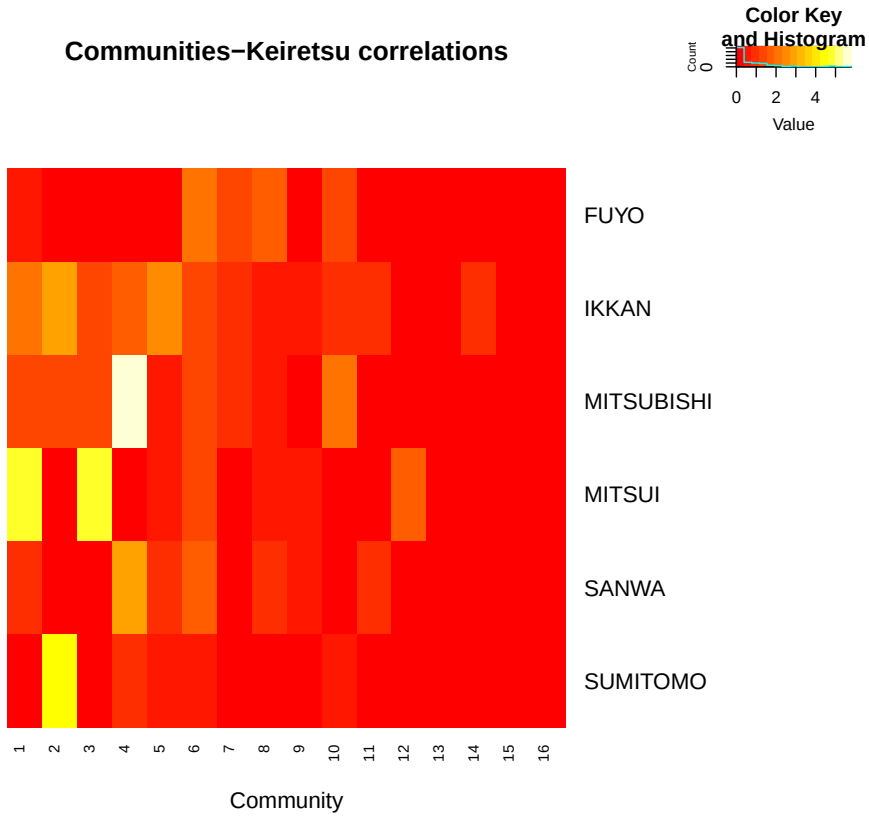
Figure 2.4 displays the bivariate joint frequencies of firms on both keiretsu and community categories as a heatmap, for the global dataset, including manufacturing firms worldwide. Hints of a non-random distribution of firms across the two categories, to be validated statistically (see sec. 2.5.2), are visible in the image. Mitsubishi- and Sumitomo-bound investors, for instance, are ostensibly concentrated in two communities.

2.5.2 Independence test

An associative measure must be put to use; since we are dealing with categorical data, the Pearson Chi-Square test, including some case-specific safeguards, have been deemed as appropriate. It has been applied to the resulting two way table, to assess whether there is any interdependence between the two attributes (this being the alternative hypothesis H_a, if the null hypothesis of independence H_0 can be rejected).

Moving on to assess the test validity, we observe that the population is fairly large, and lends itself to an in-depth analysis of subsets of interest. However, in a few specific instances (e.g. "Europe" macro-region, "Food industry" industrial sector) the number of selected observations is barely adequate. For the sake of robustness, a double line of defense has been put into

FIGURE 2.4: Bivariate joint frequencies of Japanese investors; worldwide, manufacturing sector



place: first, a key characteristic of the Louvain community detection method has been leveraged in order to increase the number of observations: being based on a stochastic algorithm, repeated runs (1.000 iterations, in this case) return slightly different partitions, which can be accumulated into a richer dataset (and averaged out, to keep from artificially inflating the sample size, biasing the test); second, the Chi-Square p-values have been computed by Monte Carlo simulation (Tate and Hyer, 1973; Bradley and Cutcomb, 1977).

2.5.3 Business ties as weighted links

The strength of business ties, in weighted networks, is measured as Cosine Similarity (Salton and McGill, 1983). CS compares the distribution of the capital invested by pairs of co-owners into subsidiaries: a perfect match (e.g. $K_{a,1} = K_{b,1}; K_{a,2} = K_{b,2}; \ldots K_{a,n} = K_{b,n}$, where $K_{PC,S}$ is the capital K invested

by the parent company PC in subsidiary S) corresponds to a $CS = 1$; as the allocation choices diverge, CS approaches zero. Technically, it is a measure of the angle between the two vectors, and its purpose is to offer a proportional representation of the connection between investing firms going beyond the binary idea of connected vs. unconnected.

2.6 Results and discussion

Timewise, it appears that, after the first, sparsely populated 1975 snapshot, a period of strong correlation (rejection of the MRH), ensues, encompassing the following decades (1985 and 1995 snapshots). In 2005, right before the mega-mergers occur, evidence starts to wane.

Geographically, an indisputable difference is observable between macro-areas with medium to strong significance (Asean countries, China and Taiwan, Northern America) and one area where there is no evidence for the persistence of HK structures at all (Europe).

When a disaggregated analysis is performed by industrial activity, the only, albeit extremely sizeable, sector with unambiguously strong correlation is "chemical"; "textile" does not offer a strong enough evidence for the existence of HK (there appears to be a very weak correlation for the unweighted network case); the independence hypothesis cannot be rejected for "food" either, since the small number of observations keeps us from achieving robust and conclusive results. Automotive yields ambiguous results (correlated when weighted, uncorrelated when unweighted); a tentative explanation, which would require a more in-depth analysis out of the scope of this chapter, may lie in the presence of numerous "smaller" partners, playing a minor role. Links implying these partners are weaker, and a more clear-cut HK structure appears when considering the most economically important subsidiaries only, connected by stronger links. The root of this discrepancy

	UNWEIGHTED			WEIGHTED		
	MRH Rej?	ChiSq	p-val	MRH Rej?	ChiSq	p-val
Overall	YES *	72.15	0.013	NO	56.18	0.080
1975	YES *	45.65	0.039	NO	56.10	0.140
1985	YES ***	61.07	0.001	YES ***	96.70	0.000
1995	YES ***	92.77	0.000	YES ***	91.64	0.000
2005	YES *	72.15	0.013	NO	56.18	0.080
ASEAN	YES *	52.68	0.030	YES ***	67.83	0.000
China and Taiwan	YES ***	77.51	0.000	YES ***	57.37	0.000
Northern America	YES *	78.96	0.011	YES ***	77.50	0.000
EU	NO	41.55	0.254	NO	49.11	0.150
Automobiles and Parts	NO	70.39	0.982	YES ***	91.10	0.000
Food	NO	45.67	0.320	NO	45.67	0.310
Chemicals and Pharma	YES ***	73.36	0.000	YES ***	63.43	0.000
Textiles	NO	23.32	0.080	NO	28.77	1.000

TABLE 2.6: Hypotheses tests results, unweighted (left) and weighted (right) columns

Note: ***$= p < .001$; **$= p < .01$; *$= p < .05$

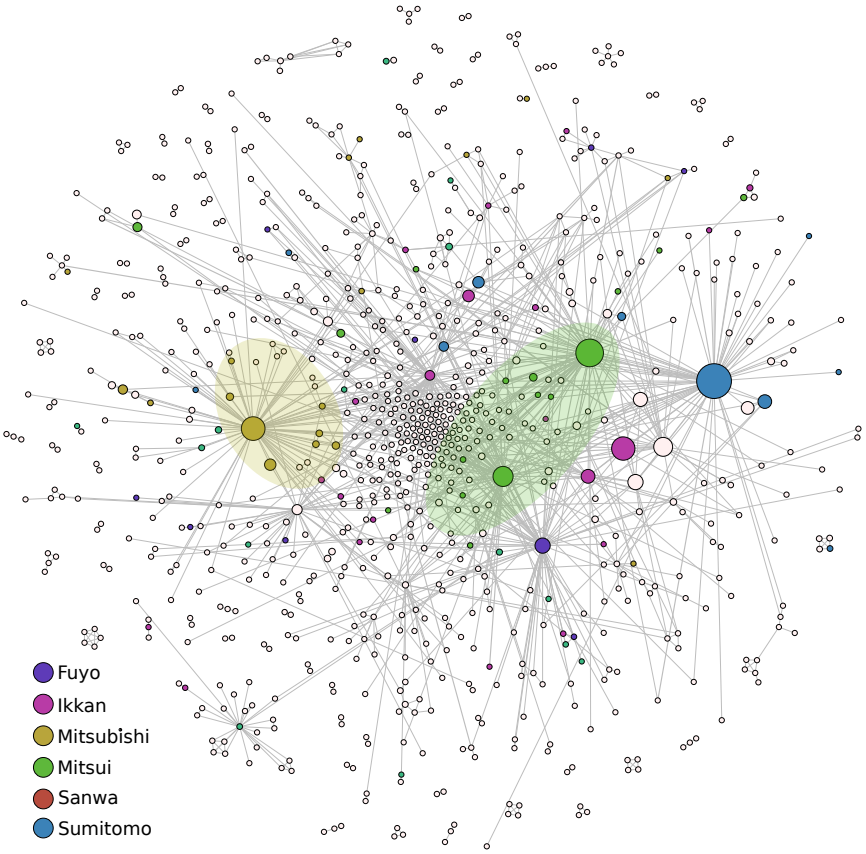
might be traced back in the bulk of smaller partners whose importance is correctly rescaled through the links' weight, thus evidencing the correlation with the HK.

In order to assess the validity of our findings, the results for every hypothesis tested have been contrasted to an alternative network Configuration Model (Newman, 2003), where the network structure is destroyed, while preserving the degree for single nodes, through a random rewiring process. Intuitively, this procedure is equivalent to blindly creating economic partnerships. In every single instance, any hint of significance disappears completely, showing p-values very close to 1.

2.7 Conclusions

In this work, we have presented an analysis of the TKZ dataset, using community detection tools, that led us to reject the MRH in the strong form.

FIGURE 2.5: Worldwide Japanese investors network: business ties and Big-6 membership. Highlighted: Mitsubishi and Mitsui clusters



Indeed, we were able to display quantitative evidence that the communities embedded in Japanese business networks strongly correlate with the HK structure described in the 2000 "Keiretsu Survey" dataset.

To give a brief summarizing overview, as a first step, we obtained from the 2006 TKZ dataset a "co-ownership graph", so that two firms are linked in our approach if they both invest in an overseas subsidiary. Through community detection algorithms, the taxonomy of high-density clusters emerged from this graph so that each firm is classified into a given cluster, as described in Sec. 2.5. This classification, purely arising from explicit business ties (the co-ownership), was then compared to the keiretsu one, described in the 2000 TKZ dataset.

Our results, summarized in Table 2.6 strongly point to a clear correlation between the intrinsic network organization and the HK, albeit with some fluctuations when one considers more specific subsets, eventually flawed by the lack of statistics, as for the investments in Europe and the food Iindustry. Another interesting point unveiled by the analysis is, in the automotive sector, the discrepancy between the clear correlation shown by the weighted network with respect to the unweighted configuration. Finally, to test the soundness of our findings, we provided, as a comparison, a null model by shuffling the links and destroying the existing correlations with the Configuration Model. This test, disrupting any network structure, leads to p values near to 1 and, thus, further proves that the Japanese business network bears a strong intrinsic mark in its organization.

Chapter 3

Revealing evolutions in dynamical networks

3.1 Introduction

The description of large temporal graphs requires effective methods giving an appropriate mesoscopic partition. Many approaches exist today to detect "communities", i.e. groups of nodes that are densely connected (Fortunato, 2010), in static graphs. However, many networks are intrinsically dynamical, and need a dynamic mesoscale description, as interpreting them as static networks would cause loss of important information (Holme and Saramäki, 2012; Holme, 2015). For example, dynamic processes such as the emergence of new scientific disciplines, their fusion, split or death need a mesoscopic description of the evolving network of scientific articles.

There are two straightforward approaches to describe an evolving network using methods developed for static networks. The first finds the community structure of the aggregated network, i.e. the network found by aggregating the nodes and their links at all times. However, this approach discards most temporal information, and may lead to inappropriate descriptions, as very different dynamic data can give rise to the identical static graphs (Berger-Wolf and Saia, 2006). To avoid this problem, the opposite approach closely

follows the evolutions and builds networks for successive time slices by selecting the relevant nodes and edges. Then, the mesoscopic structure of each of these slices is found independently and the structures are connected in various ways to obtain a temporal description (Berger-Wolf and Saia, 2006; Palla, Barabási, and Vicsek, 2007; Rosvall and Bergstrom, 2010; Chavalarias and Cointet, 2013). By using an optimal structural description at each time slice, this method avoids the inertia of the aggregated approach. Its main drawback lies in the inherent fuzziness of the communities, which leads to "noise" and artificial mesoscopic evolutions, with no counterpart in the real evolutions of the data. For example, rather different partitions have a very close modularity (Good, Montjoye, and Clauset, 2010), and minor changes in the network may lead to quite different partitions in successive time slices, which would be inadequately interpreted as major structural changes.

Several methods have been proposed to overcome the problems of these two extreme approaches (Gauvin et al., 2015; Peel and Clauset, 2015; Mucha et al., 2010; Kawadia and Sreenivasan, 2012). Here, we present a new approach that distinguishes real trends and noise in the mesoscopic description of social data using the continuity of social evolutions. To be able to follow the dynamics, we compute partitions for each time slice, but to avoid transients generated by noise, we modify the community description at time t using the structures found at times $t-1$ and $t+1$. We show the relevance of our method on the analysis of a scientific network showing the birth of a new subfield, wavelet analysis. This field represents a difficult test because it has arisen out of the collaboration of several disciplines, producing a rich history, made by many entangled streams.

3.2 Method description

Our method consists in four steps:

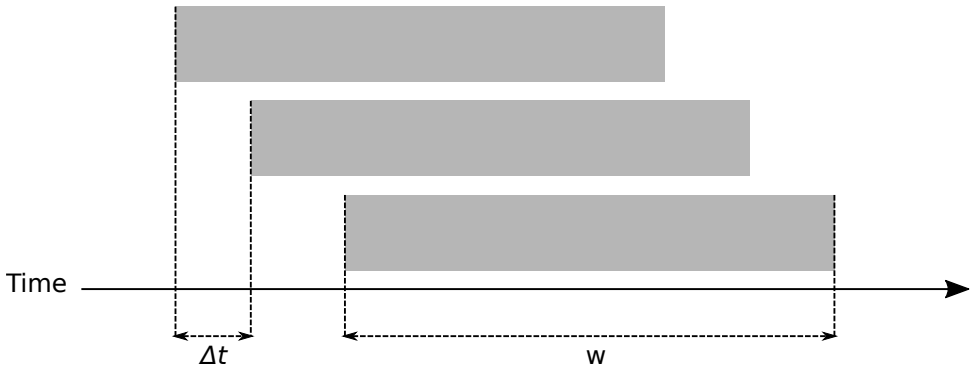


FIGURE 3.1: Sliding temporal windows

1. The dataset is first divided into temporal windows of w years, translated by Δt years (Fig. 3.1).

2. In the second step, community detection is carried out independently for each window by any method. This leads to a structure that follows as closely as possible the real mesoscale dynamics, at the price of some noise. To selectively delete the noise, while keeping the real evolutions, one has to split or merge communities at each slice, depending on the relations between the successive communities on longer time scales.

3. For this, the third step systematically computes all the similarities between communities at times $t-2$, $t-1$, t, $t+1$ and $t+2$. For each community at time t, we keep only the most similar communities at times $t-2$ through $t+2$, thus defining its "ancestor" (most similar community at $t-2$) P_{t-2}, "predecessor" (at $t-1$) P_{t-1}, "successor" (at $t+1$) S_{t+1}, and "grandchild" (at $t+2$) S_{t+2}. These strong long-term links allow to discriminate real evolutions from noise, by taking advantage of the relative continuity and stability of social evolutions on appropriate time scales. For example, and with regard to the dataset used, a new scientific field does not appear and disappear in a single year.

4. The fourth (and final) step then uses this long-term information to iteratively select all the time windows and optimize the complexity score

(Equation 1). For this, we merge communities that appear to be unduly split by the independent community detection (Figure 2a), and split communities that appear to be artificial merges (Figure 2c). In practice, we identify artificial merges at time t by the links between the "predecessor" communities (at t-1) and the "successor" ones (at t+1). If these two are linked (as in Figure 2c), then we assume that these two trends represent the real evolution, and the merge at time t arises out of noise in the community detection. We then split the community, attributing the nodes to each of the trends by a simple intersection procedure (for details, see SI, par. 1). In any other case, when there are missing links between the communities (as in Figure 2d), we assume that a real merge has been detected, which is then followed by a split between two different streams. The same procedure is applied to distinguish between real and artificial splits (Figure 2a-b). This procedure goes on as long as there exists an artificial split or merge.

5. At the end of the procedure, we obtain a description of the network evolution at the mesoscale, the unit of description being now several streams of connected communities. Note, however, that the final description may depend on the set of initial partitions. To render our method robust, we compute a "complexity" score (Equation 1) for different final descriptions and use the one with the highest score, leading to the "richer" story that can be told avoiding noise. The merit of our approach is, by eliminating most of the noise, to limit these complex turbulent regions to the real transformations that should not be discarded: things should be made simple, but not too simple. This score is computed as (eq. 3.1):

$$C_s = \frac{\sum_{u \in (u_s \cup u_m)} s_u - \sum_{u \in (u_r \cup u_x)} s_u}{\sum_{u \in G} s_u} \tag{3.1}$$

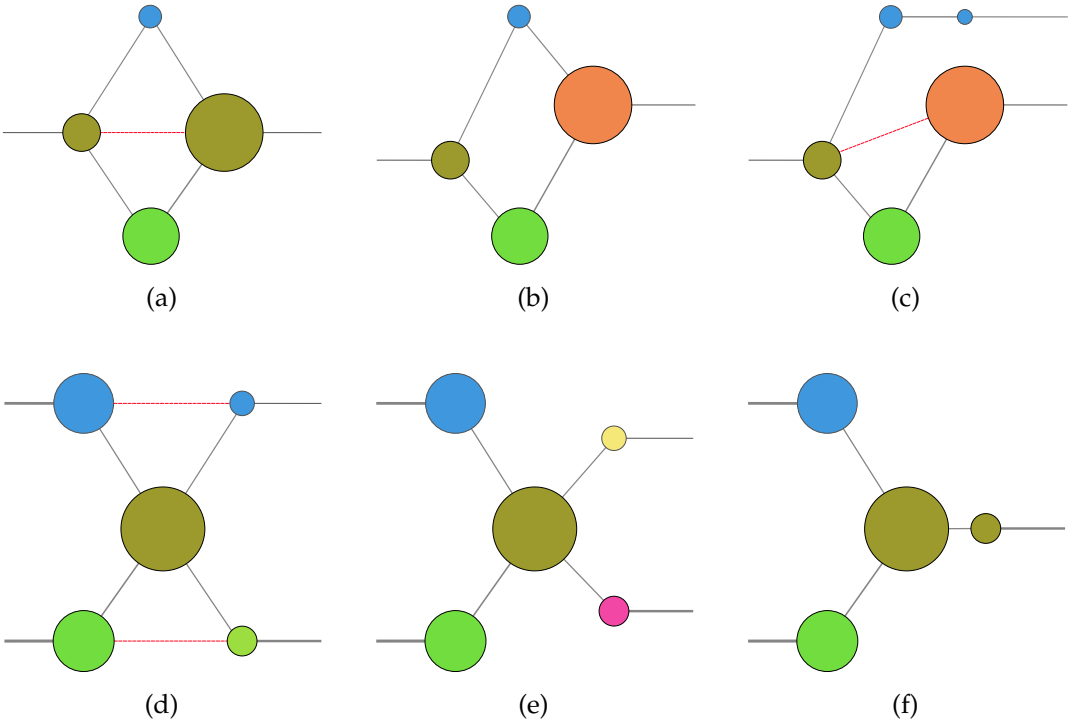


FIGURE 3.2: Ephemeral (artificial) versus structural (real) events

Where:

$u =$ nodes in G

$s_u =$ size of node u

$u_s =$ resulting split nodes in "structural" (real) splits

$u_m =$ resulting merged node in "structural" (real) merges

$u_r =$ resulting split nodes in ephemeral (noise) splits

$u_x =$ resulting merged node in ephemeral (noise) merges

To visualize the output, we align the communities that belong to the same "laminar stream", defined as a succession of communities that are all connected, by both $t \pm 1$ and $t \pm 2$ links. More formally, a laminar stream LS is defined as an ensemble of communities C_i such as:

$$C_i \in LS \Leftrightarrow P_{t-2}(C_i) \in LS \wedge P_{t-1}(C_i) \in LS \wedge S_{t+1}(C_i) \in LS \wedge S_{t+2}(C_i) \in LS \tag{3.2}$$

where $P_{t-1}(C_i)$ is the predecessor of C_i at $t-1$, etc. In general, real systems are not only composed of laminar streams: there are some "turbulent regions", where real splits produce new streams, flows become intermingled and new subfields are generated. This turns out to be the case in the real case application we describe below.

3.3 Emergence and evolution of a new scientific field: wavelet analysis

We test the method on an evolving network of scientific articles related to the emergence of a new field: wavelets analysis. This technique, developed through collaborations among mathematicians, physicists and electrical engineers, has been fundamental for signal/image processing, leading for example to the well-known jpeg compression format. Wavelets history is interesting as a test case because it is a recently born subfield (seminal paper in 1984), for which robust scientometrics records are available. To define the relevant set of publications, we identified 83 key actors of the early developments of the field (see A.5). The list was established using expert advice (one of the authors, PF) and bibliographic searches. We then retrieved all their publications (from 1970 to 2012), obtaining 6,500 records from Web of Science. We used 4-years wide time slices ($w = 4$), separated by one year ($\Delta t = 1$). For each slice, we first defined a network using the articles as nodes, linked by their common references (bibliographic coupling, Kessler (1963)), articles sharing less than 2 references are not linked). We then follow the method described above, using maximization of modularity for each



FIGURE 3.3: Overview of the history of wavelets

Note: streams are labelled according to the subfield within wavelets development (see text for details). Each community is positioned according to year (x-axis), aligned according to streams (y-axis) Zoomable version available at http://perso.ens-lyon.fr/matteo.morini/wavelets/flows/disciplines.pdf

slice and the Jaccard similarity index (Jaccard, 1901) to compute the similarities between the successive communities. The final result is represented in Figure 3.3.

We can now address two important points:

1. What have we learnt about wavelets evolutions using our method?

2. Methodological: what do we learn about our method from this example? How important are artificial splits/merges, quantitatively and qualitatively, i.e. to understand the history of wavelets?

3.4 Automatic wavelets history

A first idea about wavelets evolutions can be derived from the evolution of modularity (Fig. 3.5). Roughly speaking, a high modularity value corresponds to isolated clusters, while low values point to highly interconnected networks. The analysis shows that there are three main stages. In an initial phase (before 1985), researchers work in different, quite unrelated fields and modularity is high (the network of all articles in this period is shown in Fig. S5). Then, in the 1990s, wavelets appear as a common topic whose use gains momentum, defining a new, specific field that interlinks the publications of our set of authors, leading to a minimum in modularity. After this, modularity increases again, pointing to a new, softer divergence, as the initial levels are not reached. Wavelets become a mature tool, that are less an object of interest per se, serving instead a more ancillary role within specialized communities and paving the way for new avenues of research, by developing new tools (as "compressed sensing") or applying wavelets to specific domains, such as Astrophysics images.

Our approach reveals the major structural flows that define the subfields within wavelets development (Fig. 3.3). For each stream, we indicate its name, the main author and the initial/ending dates.

- The stream "Foundations of wavelets" (1983-2006) starts in 1983 with foundational articles by mathematician Alexis Grossmann. Most wavelets research streams emerge from it, as "Time-frequency analysis" (Flandrin, 1989-2009) and "Component separation" (wavelets without orthonormal bases, Szu, 1989-2009). Starting from mathematical physics, this streams builds wavelets as a rigorous mathematical formalism (80% of its articles are published in Mathematics journals), but adapted to Engineering concerns.

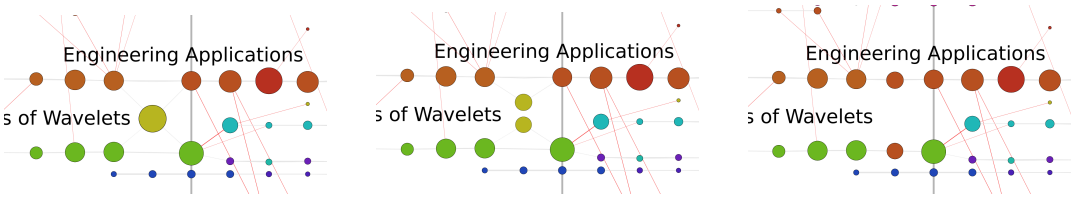
- In 1990, a central stream, the stem from which most of the subsequent streams will emerge ("Engineering applications", 1990-2009), is created by the fusion of Vetterli's research with a split of the founding stream. This subfield is less concerned by theoretical developments than by practical applications, and most of its articles are published in Engineering journals. The stream "Representation" (Unser, 1986-1992) joins it in 1993, leading to a focus on design. The most important subfields originating in "Applications" are:
  - "Inverse problems & sparsity in image analysis" (Starck, 1993-2009), which after focusing on applications on astrophysics images, deals with more general problems in image analysis. It will lead to another important stream, "Compressed sensing" (Baraniuk, 2004-2009)
  - "Structural models" (Wilsky, 1994-2009)
  - "Image coding", building the theoretical foundations of image coding (Vetterli, 1998-2006)
  - "Data hiding" (Ramchandran, 1998-2009)
- Note that there are also some "laminar flows", that interact only peripherally with other lines of research, leading to a linear, simple sequence of communities. Examples of these relatively independent lines of research are the group lead by Alain Arneodo ("Multifractal", 1985-2009), "Frames" (Gröchenig, 2000-2009), "Video quality" (Bovik, 1999-2009). These laminar flows represent subfields that apply wavelets to specific objects, without contributing much to the methodological developments.

Finally, it is instructive to look for the position of Yves Meyer, the 2017 prestigious Abel prize for "his pivotal role in the development of the mathematical theory of wavelets". As the number of his publications is not very high, he does not appear explicitly as the main author of any stream. However, his publications are highly cited in the stream "foundations of wavelets", revealing his importance for the mathematical developments. His "pivotal role" of connecting ideas and people, notably in conferences, cannot be seen in our network only made from publications.

3.5 Test of the method

Overall, our method has lead to the split (12) and merge (12) of 24 communities, representing 10% of all the articles in the database. An example of an important artificial merge detected by our method, similar to the one sketched in Fig. 3.2(d), is given in Fig. 3.4. Even if it is clear, looking at the overall history, that there existed two distinct streams of research for 20 years, the independent initial partition merged the communities from these two trends in 1994. Since there exist links (i.e. shared references) among the articles of the two communities (see the articles' network in Fig. S4), there is a significant probability that an independent partition algorithm will gather them in a single community. Our method allows to avoid this artificial merge of two distinct streams of research, which belong to different disciplinary traditions, as one subfield is focused on Mathematics, while the other privileges Engineering.

More generally, Fig. 3.5 shows that the rearrangements of partitions demanded by the maintenance of the streams flows leads to negligible losses in the quality of the instantaneous partitions as quantified by modularity. This is important, as it shows that we maintain a close adaptation to the temporal variations, while choosing the partition that best fits the overall evolution.



(a) (b) (c)

FIGURE 3.4: Detail of 1994 split. From left to right, the unduly merged communities are split and assigned to their respective streams.

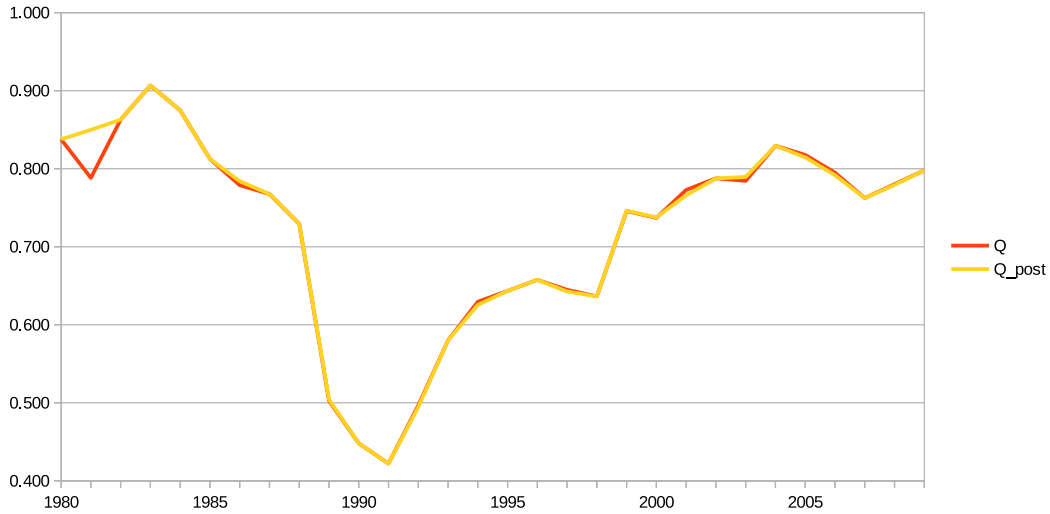


FIGURE 3.5: Modularity evolution, pre- and post-filtering.

There is a clear minimum in modularity for the year 1991, pointing to a homogeneous network without much structure, which results from the effective mixing of different disciplinary traditions around the new object (wavelets).

3.6 Comparison to other methods

Our approach offers decisive advantages over existing methods:

- It handles naturally networks in which nodes appear or disappear at each time step, which is impossible or cumbersome for other methods.
- Contrary to generative models (Jacobs and Clauset, 2014; Peixoto and Rosvall, 2016; Xu, Kliger, and Hero III, 2014), we do not need to define

an a priori network community structure (for example, a block model) which may be unadapted to the data.

- Claveau and Gingras (2016) have studied the history of economics using scientometrics data and an approach quite similar to ours. However, to determine the partition at time t, they initialize the Louvain algorithm with the partition obtained for the preceding time step. Their approach is therefore limited to partitioning by this algorithm. Moreover, the authors do not justify why their approach represents a sound way of adding some inertia to real-time partitioning.

3.7 Concluding remarks and future work

To make sense of transformations, we need evolving categories that can, at the same time, readily adapt to the changes and maintain the continuity of the description. Our method starts from the idea that the unity of an evolving social process rests on the continuity of its transformations, and uses the available mid-term temporal information to reveal structural trends from noisy data, without the assumption of an a priori community structure. It can be adapted to any partitioning method and to any similarity measure between communities at different times. Used on scientific data, our method automatically produces a rich historical account, an objective raw material to be discussed by science historians.

There is much room for improvement. The relevant time scales (w, Δt) have to be chosen from expert knowledge, and we cannot deal with real-time data, as we use the future to infer the best present partition. We now work to introduce, through a hidden Markov model, an explicit meso temporal scale at which transformations (splits/merges) are supposed to happen for a pair of streams.

Appendix A

Supplementary Information

A.1 Modified Brandes algorithm

Bridgeness algorithm, inspired by Brandes' "faster algorithm" (Brandes, 2001)

SP[s,t] ←precompute all shortest distances matrix/dictionary
CB[v] ← 0, v ∈ V ;
for s ∈ V **do**
 S ← empty stack;
 P[w] ← empty list, w ∈ V ;
 σ[t] ← 0, t ∈ V ; σ[s] ← 1;
 d[t] ← −1, t ∈ V ; d[s] ← 0;
 Q ← empty queue;
 enqueue s → Q;
 while Q not empty **do**
  dequeue v ← Q;
  push v → S;
  foreach neighbor w of v **do**
   // w found for the first time?
   if d[w] < 0 **then**
    enqueue w → Q;
    d[w] ← d[v] + 1;

```
    end
    // shortest path to w via v?
    if d[w] = d[v] + 1 then
     σ[w] ← σ[w] + σ[v];
     append v → P[w];
    end
   end
  end
  δ[v] ← 0, v ∈ V ;
  // S returns vertices in order of non-increasing distance from s
  while S not empty do
   pop w ← S;
   for v ∈ P[w] do δ[v] ← δ[v] + σ[v]/σ[w] · (1 + δ[w]);
   if SP[w,s]>1 then CB[w] ← CB[w] + δ[w];
  end
 end
```

A.2 Case study on a synthetic network community

The specificity of bridgeness and the influence of the degree, which prevents BC from identifying correctly the most important bridges, can be exemplified by examining the scores of nodes in cluster 5 of the synthetic network. This cluster is linked to cluster 13 by 5 connections (through nodes 248, 861, 471, 576 and 758) and to cluster 1 by a single connection (through node 232). BC gives roughly the same score to nodes 232 and 248, while bridgeness attributes a score almost 4 times higher to node 232, correctly pointing out the importance of this single bridge between clusters 5 and 1. This is because BC is confused by the high degree of node 248 (41) as compared to node 232 low degree (20). Therefore, by counting all the shortest paths, BC attributes

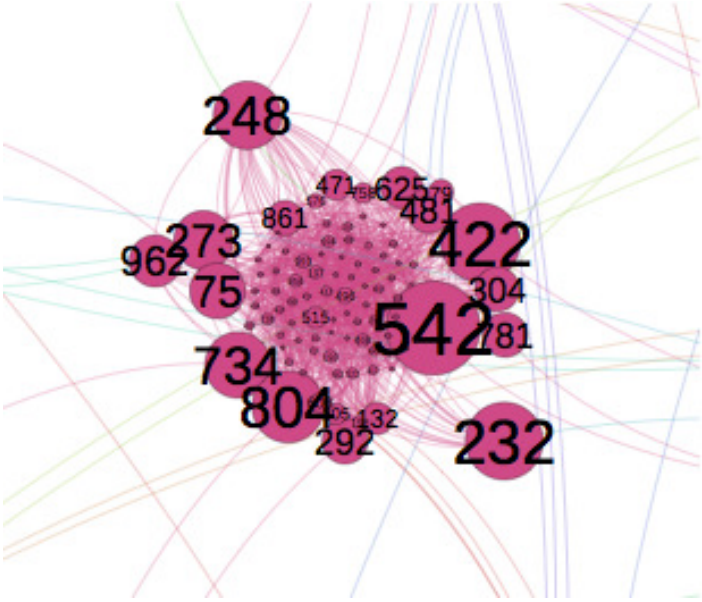


FIGURE S1: Zoom on cluster 5 of the synthetic network. The numbers represent labels, while the size of the nodes is proportional to their BC score.

too high a bridging score to node 248. Second problem with BC, it gives a high score to nodes that are not connected to other communities, merely because they are local centers, *i.e.* they have a high degree. For example, node 515 obtains a higher BC score than node 758 (Table S1), even if node 515 has no connection to other communities (but degree 49), contrary to node 758 (connected to cluster 5, but degree 23). Bridgeness never ranks higher local centers than global bridges: here, it correctly assigns a 5 times higher score to node 758 than to node 515.

A.3 Japanese co-investment network facts

The global network, complete with data from every industrial sector, once partitioned in communities (keeping into account the links weight), is composed by clearly defined clusters (as expected, given the high modularity value shown earlier), the most important five, containing at least 5% of the total number of nodes, total almost half of the nodes (298, or 49.42%).

TABLE S1: Nodes in community 5 of the synthetic network, ranked by decreasing BC (see text)

Id	Stirling	Modularity Class	Betweenness	Bridgeness	Degree
542	0.0222	5	9173.71	2644.62	44
422	0.0278	5	7714.27	3855.62	35
232	0.0950	5	7551.22	5846.86	20
804	0.0285	5	6995.63	2824.64	34
248	0.0082	5	6588.65	1624.30	48
734	0.0907	5	6410.31	4373.72	21
273	0.0322	5	5698.28	2631.59	30
75	0.0868	5	5349.47	3558.31	22
962	0.0399	5	4989.66	2951.45	24
292	0.0399	5	4377.77	1939.06	24
481	0.0256	5	4305.68	1796.92	25
781	0.0475	5	4257.93	2200.21	20
304	0.0434	5	4221.64	2467.65	22
625	0.0202	5	3964.21	1314.62	32
861	0.0108	5	3295.01	714.44	36
132	0.0200	5	2985.45	1157.49	24
471	0.0154	5	2865.07	1296.38	25
79	0.0302	5	2256.02	1004.28	21
205	0.0208	5	1921.65	788.51	23
515	0.0000	5	1884.07	86.45	49
758	0.0166	5	1791.80	435.66	23
608	0.0200	5	1777.54	522.75	24

Each node has a "size" representing the total shares worth (expressed in 2009 constant US$) owned by the investor.

Graph metrics

- # Nodes: 603
- # Edges: 972
- Average Degree: 3.224
- Average Weighted Degree: 0.232
- Graph Density: $5.0\mathrm{e}{-3}$
- Modularity: 0.93

- Average Path Length: 3.92

Main communities and salient nodes A few salient firms are selected from each community, and identified by their TKZ code.

id 19 97 nodes (16.09%) 805300 803100 540300

(Sumitomo Corp.; Mitsui and Co. Ltd.; JFE Steel Corporation)

id 32 77 nodes (12.77%) 805800

(Mitsubishi Corp.)

id 62 46 nodes (7.63%) 800100 400500

(Itochu Corp.; Sumitomo Chemical Co. Ltd.)

id 53 46 nodes (7.63%) 801500

(Toyota Tsusho Corp.)

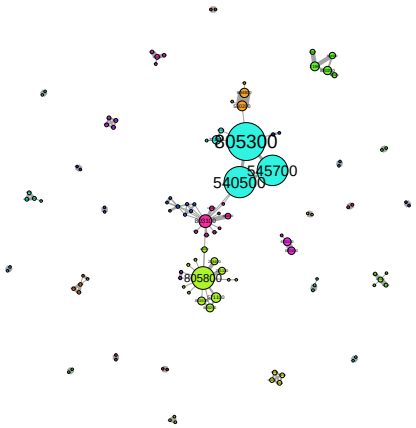
id 35 32 nodes (5.13%) 800200

(Marubeni Corp.)

A.3.1 A small digression on Bridgeness Centrality (Jensen et al., 2016):

Looking for nodes acting as bridges between distant communities, 520100 (Asahi Glass Co., Ltd., Mitsubishi group, Sanwa Keiretsu) appears the most interesting, being sited between the central part of the global network and the small cluster organized around 808800 (Iwatani International Corp., a trading company investing mainly in China, but also Singapore, Thailand, Malaysia, USA, Australia and Germany), including gas and other goods.

FIGURE S2: Northern America (incl. U.S. and Canada) network



A.4 Geographical analysis

A.4.1 Northern America (incl. U.S. and Canada)

The Northern-American network, composed of 112 nodes, is disconnected (44 nodes only belong to the main connected component). Community 7 only includes 6 nodes, but includes the three the top (by worth) enterprises, all belonging to Sumitomo Corporation; community 15 is detached from the main component.

Graph metrics

- # Nodes: 112
- # Edges: 116
- Average Degree: 2.071
- Average Weighted Degree: 0.400
- Graph Density: 1.9e−2

- Modularity: 0.95
- Average Path Length: 3.62

Main communities and salient nodes

id 14 13 nodes (11.61%) 805800 571100 831500 840200

(Mitsubishi Corp.; Mitsubishi Materials Corp.; Bank of Tokyo–Mitsubishi UFJ, Ltd.; Mitsubishi UFJ Trust and Banking Corporation)

id 16 9 nodes (8.04%) 803100 404100 631300 542300

(Mitsui and Co. Ltd.; Nippon Soda Co, Ltd.; Kioritz Corp. *(agricultural machinery)*; Tokyo Steel Mfg.Co.,Ltd.)

id 13 8 nodes (7.14%) 801200 418300 787100

(Nagase & Co., Ltd.; Mitsui Chemicals, Inc.; Fukuvi Chemical Industry Co. Ltd.)

id 7 6 nodes (5.36%) 805300 545700 540500

(Sumitomo Corp.; Sumitomo Pipe & Tube Co., Ltd.; Sumitomo Metal Industries, Ltd.)

id 15 5 nodes (4.46%) 800200 2196 594900

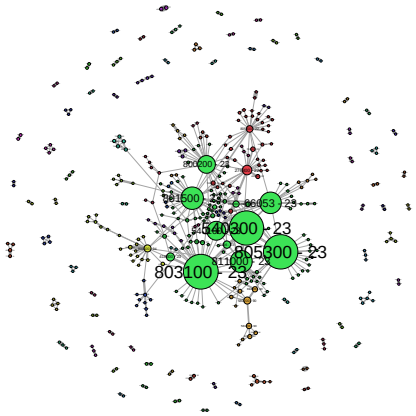
(Marubeni Corp.; Marubeni-Itochu Steel Inc.; Unipres Corp.)

A.4.2 ASEAN Countries

The Southest Asian Nations network, composed of 415 nodes, is disconnected (240 nodes belong to the main connected component).

The first and second largest communities include almost one third of the nodes.

FIGURE S3: ASEAN Network



Graph metrics

- # Nodes: 415
- # Edges: 468
- Average Degree: 2.255
- Average Weighted Degree: 0.340
- Graph Density: 5.0e−3
- Modularity: 0.97
- Average Path Length: 3.79

Main communities and salient nodes

id 23 95 nodes (22.89%) 540300 803100 805300 801500 811000

(JFE Steel Corp.; Mitsui & Co., Ltd.; Sumitomo Corp.; Toyota Tsusho Corp.; JFE Shoji Trade Corp.)

id 22 36 nodes (8.67%) 276800 805800 721100 701200

(Sojitz Corp.; Mitsubishi Corp.; Mitsubishi Motors Corp.; Kawasaki Heavy Industries, Ltd.)

id 94 14 nodes (3.37%) 800100 12118 260700

(Itochu Corp.; Kyoei Steel Ltd.; Fuji Oil Co., Ltd.)

A.4.3 China

The Chinese network, composed of 466 nodes, is disconnected (256 nodes belong to the main connected component).

Community 16: steel;

Graph metrics

- # Nodes: 466
- # Edges: 534
- Average Degree: 2.292
- Average Weighted Degree: 0.350
- Graph Density: $5.0\mathrm{e}{-3}$
- Modularity: 0.98
- Average Path Length: 4.35

Main communities and salient nodes

id 16 39 nodes (8.37%) 803100 805300 807800 540700 540100

(Mitsui & Co., Ltd; Sumitomo Corp.; Hanwa Co., Ltd.; Nisshin Steel Co., Ltd.; Nippon Steel Corp.)

id 96 38 nodes (8.15%) 800100 as Sōgō shōsha; the rest are minor, 260700, 701200, 359700...

(Itochu Corp.; Fuji Oil Co., Ltd.; Kawasaki Heavy Industries, Ltd.; Jichodo Co., Ltd.)

A.5 List of WLTC selected authors

Aldroubi, A; Antoine, JP; Antonini, M; Argoul, F; Arneodo, A; Auscher, P; Bacry, E; Baraniuk, RG; Barlaud, M; Battle, GA; Benveniste, A; Beylkin, G; Bijaoui, A; Boudreaux-Bartels, GF; Bovik, AC; Burrus, CS; Chui, CK; Cohen, A; Coifman, RR; Compo, GP; Dahmen, W; Daubechies, I; Devore, RA; Donoho, DL; Elezgaray, J; Escudie, B; Farge, M; Feauveau, JC; Feichtinger, HG; Flandrin, P; Frisch, U; Froment, J; Gopinath, RA; Grochenig, K; Grossmann, A; Guillemain, P; Haar, A; Healy, DM; Heil, C; Herley, C; Holschneider, M; Jaffard, S; Jawerth, B; Johnstone, IM; Jones, DL; Kadambe, S; Kaiser, JF; Kerkyacharian, G; Kronland-Martinet, R; Lawton, W; Lemarie-Rieusset, PG; Lu, J; Lucier, BJ; Mallat, S; Mendlovic, D; Meyer, Y; Micchelli, CA; Morlet, J; Murenzi, R; Muschietti, MA; Muzy, JF; Paul, T; Perrier, V; Picard, D; Ramchandran, K; Resnikoff, HL; Rioul, O; Saracco, G; Shao, XG; Shapiro, JM; Slezak, E; Starck, JL; Szu, HH; Tchamitchian, P; Torresani, B; Unser, M; Vaienti, S; Vetterli, M; Walter, GG; Weaver, JB; Wickerhauser, MV; Willsky, AS; Wornell, GW.

A.6 Re-splitting unduly merged communities

Occasionally, and because of the inherently noisy community detection process, groups of nodes (articles) of a time window can assigned to either two distinct communities or a single, larger community. When we weigh in the

additional temporal information from $t-1$ and $t+1$, and observe that the ambiguity can be resolved, for example when two consistently distinct streams have been merged for a single step. We then assume that the unduly merged communities have to be split back. In order to preserve continuity, each article belonging to the wrongly merged community C_0 is assigned to one of two new communities, C_{01}, C_{02}. We define two sets of articles, U_a and U_b, one for each of the two streams, S_a and S_b, which correspond to the union of nodes appearing within each couple of predecessor/successor: $P_{a,t-1}$ and $P_{a,t+1}$, and $P_{b,t-1}$ and $P_{b,t+1}$ respectively. Nodes from C_0 which belong to the set U_a are assigned to C_{01}; similarly, nodes belonging to U_b end up into C_{02}. Because of the fuzziness of communities, a node can appear both in $P_{a,t-1}$ and $P_{b,t+1}$; in this case, we assign it to the stream to which it is connected more strongly (or randomly if both weights are equal).

A.7 Network structure examples, colors according to community

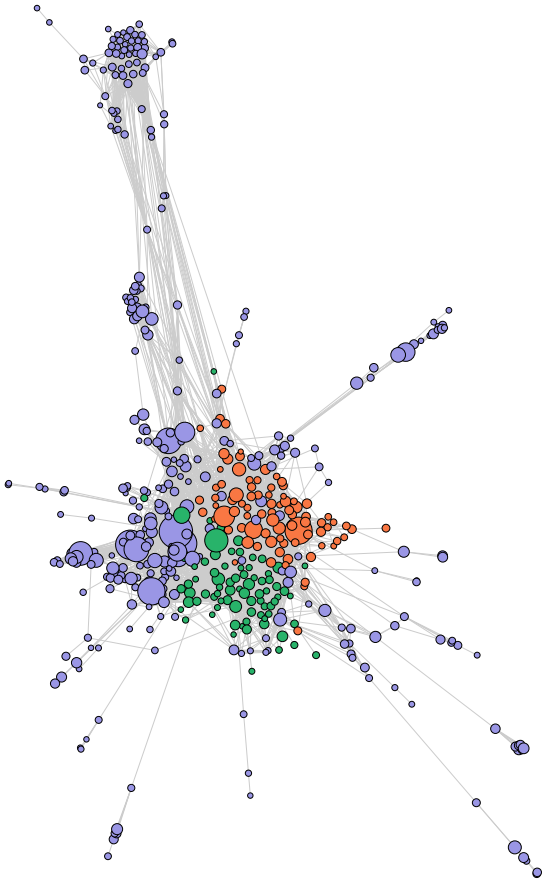


FIGURE S4: Year 1994: an ambiguous group of nodes, can be either split in two or merged in one single community, colored in green and orange; see also main text, section 3.5



FIGURE S5: Full articles network, time window 1994–1997

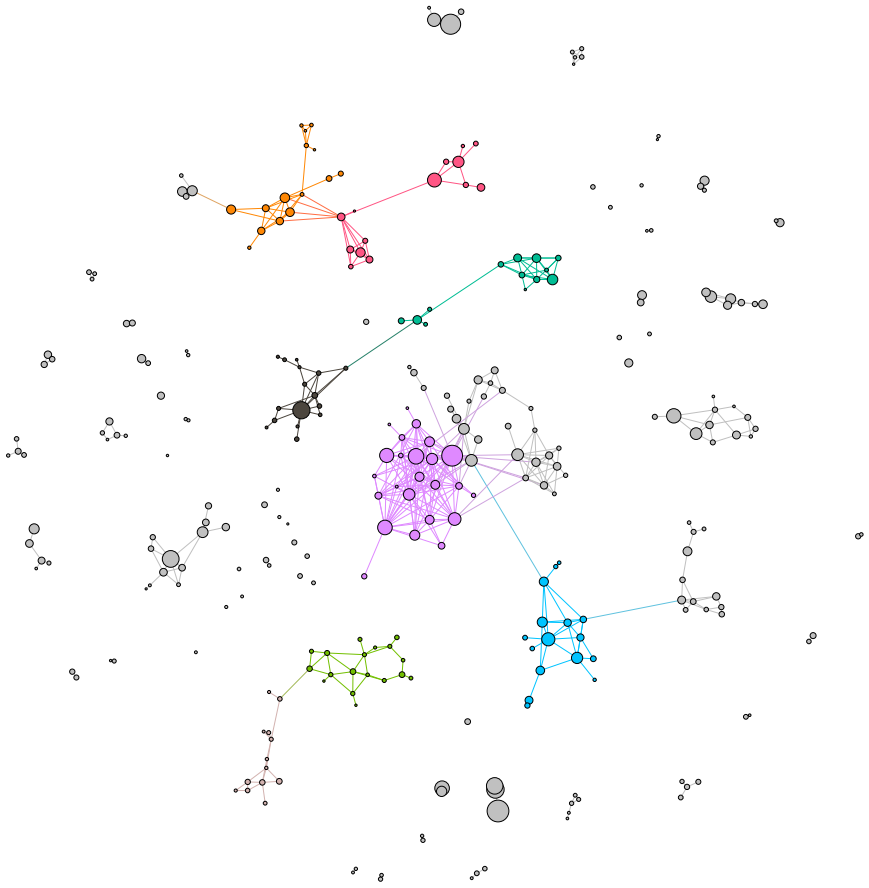


FIGURE S6: Full articles network, up to year 1982

Bibliography

Aoki, M. and G. Saxonhouse, eds. (2000). *Finance, governance, and competitiveness in Japan*. Oxford University Press.

Aoki, M. (1984a). *The co-operative game theory of the firm*. Oxford University Press.

— (1984b). *The economic analysis of the Japanese firm*. North-Holland.

Bassino, J.-P. and J. G. Williamson (2015). *From commodity booms to economic miracles: why Southeast Asian industry lagged behind*. Discussion Paper 10611. CEPR. URL: http://ssrn.com/abstract=2610726.

Bastian, M., S. Heymann, and M. Jacomy (2009). "Gephi: An Open Source Software for Exploring and Manipulating Networks". In: *International AAAI Conference on Weblogs and Social Media*.

Belderbos, R. and M. Carree (2002). "The location of Japanese investments in China: Agglomeration effects, keiretsu, and firm heterogeneity". In: *Journal of the Japanese and International Economies* 16.2, pp. 194–211.

Belderbos, R. and L. Sleuwaegen (1996). "Japanese firms and the decision to invest abroad: business groups and regional core networks". In: *The review of Economics and Statistics*, pp. 214–220.

Berger-Wolf, T. Y. and J. Saia (2006). "A framework for analysis of dynamic social networks". In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 523–528.

Blondel, V. D. et al. (2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008.

Bonacich, P. (1987). "Power and centrality: A family of measures". In: *American journal of sociology* 92.5, pp. 1170–1182.

Borgatti, S. P. (2005). "Centrality and network flow". In: *Social networks* 27.1, pp. 55–71.

Bradley, D. R. and S. Cutcomb (1977). "Monte Carlo simulations and the chi-square test of independence". In: *Behavior Research Methods* 9.2, pp. 193–201.

Brandes, U. (2001). "A faster algorithm for betweenness centrality". In: *Journal of Mathematical Sociology* 25.2, pp. 163–177.

Burt, R. S. (2005). *Brokerage and closure: An introduction to social capital*. Oxford: Oxford University Press.

Chandler Jr, A. D. (1962). *Strategy and Structure: Chapters in the History of the Industrial Enterprise*. MIT press.

— (1977). *The visible hand: The managerial revolution in American business*. Harvard University Press.

Chavalarias, D. and J.-P. Cointet (2013). "Phylomemetic patterns in science evolution—the rise and fall of scientific fields". In: *PloS one* 8.2, e54847.

Cheng, X.-Q. et al. (2010). "Bridgeness: a local index on edge significance in maintaining global connectivity". In: *Journal of Statistical Mechanics: Theory and Experiment* 2010.10, P10011.

Claveau, F. and Y. Gingras (2016). "Macrodynamics of economics: A bibliometric history". In: *History of Political Economy* 48.4, pp. 551–592.

Coser, R. (1975). "The Complexity of Roles as Seedbed of Individual Autonomy." In: *The Idea of Social Structure: Essays in Honor of Robert Merton.* Ed. by L. Coser. New York: Harcourt Brace Jovanovich.

David, T. and G. Westerhuis (2014). *The power of corporate networks: A comparative and historical perspective*. Vol. 26. Routledge.

Delios, A. and W. I. Henisz (2000). "Japanese firms' investment strategies in emerging economies". In: *Academy of Management journal* 43.3, pp. 305–323.

Estrada, E. and J. A. Rodriguez-Velazquez (2005). "Subgraph centrality in complex networks". In: *Physical Review E* 71.5, p. 056103.

Etzkowitz, H. and L. Leydesdorff (2000). "The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations". In: *Research policy* 29.2, pp. 109–123.

Euler, L. (1736). "Solutio problematis ad geometriam situs pertinentis". In: *Commentarii academiae scientiarum Petropolitanae* 8, pp. 128–140.

Forsgren, M. and J. Johanson (1992). *Managing networks in international business*. Routledge.

Fortunato, S. (2010). "Community detection in graphs". In: *Physics reports* 486.3, pp. 75–174.

Fortunato, S. and M. Barthelemy (2007). "Resolution limit in community detection". In: *Proceedings of the National Academy of Sciences* 104.1, pp. 36–41.

Freeman, L. C. (1977). "A set of measures of centrality based on betweenness". In: *Sociometry*, pp. 35–41.

— (1979). "Centrality in social networks conceptual clarification". In: *Social networks* 1.3, pp. 215–239.

Gauvin, L. et al. (2015). "Revealing latent factors of temporal networks for mesoscale intervention in epidemic spread". In: *preprint arXiv:1501.02758*.

Gerlach, M. L. (1992). *Alliance capitalism: The social organization of Japanese business*. Berkeley: University of California Press.

Goh, K.-I. et al. (2003). "Betweenness centrality correlation in social networks". In: *Physical Review E* 67.1, p. 017101.

Good, B. H., Y.-A. de Montjoye, and A. Clauset (2010). "Performance of modularity maximization in practical contexts". In: *Physical Review E* 81.4, p. 046106.

Granovetter, M. S. (1973). "The strength of weak ties". In: *American journal of sociology* 78.6, pp. 1360–1380.

Grauwin, S. and P. Jensen (2011). "Mapping scientific institutions". In: *Scientometrics* 89.3, pp. 943–954.

Hoang, H. and A. Yi (2015). "Network-based research in entrepreneurship: A decade in review". In: *Foundations and Trends® in Entrepreneurship* 11.1, pp. 1–54.

Holme, P. (2015). "Modern temporal network theory: a colloquium". In: *The European Physical Journal B* 88.9, p. 234.

Holme, P. and J. Saramäki (2012). "Temporal networks". In: *Physics reports* 519.3, pp. 97–125.

Hric, D., R. K. Darst, and S. Fortunato (2014). "Community detection in networks: Structural communities versus ground truth". In: *Physical Review E* 90.6, p. 062805.

Jaccard, P. (1901). "Étude comparative de la distribution florale dans une portion des Alpes et des Jura". In: *Bulletin de la Société vaudoise des Sciences Naturelles* 37, pp. 547–579.

Jacobs, A. Z. and A. Clauset (2014). "A unified view of generative models for networks: models, methods, opportunities, and challenges". In: *preprint arXiv:1411.4070*.

Jensen, P. and K. Lutkouskaya (2014). "The many dimensions of laboratories' interdisciplinarity". In: *Scientometrics* 98.1, pp. 619–631.

Jensen, P. et al. (2016). "Detecting global bridges in networks". In: *Journal of Complex Networks* 4.3, pp. 319–329.

Kawadia, V. and S. Sreenivasan (2012). "Sequential detection of temporal communities by estrangement confinement". In: *Scientific reports* 2, p. 794.

Kessler, M. M. (1963). "Bibliographic coupling between scientific papers". In: *Journal of the Association for Information Science and Technology* 14.1, pp. 10–25.

Kilkenny, M. and N. Fuller-Love (2014). "Network analysis and business networks". In: *International Journal of Entrepreneurship and Small Business 15* 21.3, pp. 303–316.

Koibuchi, S. and T. Okazaki (2014). "Evolution of Corporate Networks in Twentieth Century Japan". In: *The Power of Corporate Networks: A Comparative and Historical Perspective*. Ed. by T. David and G. Westerhuis. Routledge, pp. 296–304.

Kono, T. (1984). *Strategy and structure of Japanese enterprises*. Routledge.

Lancichinetti, A., S. Fortunato, and F. Radicchi (2008). "Benchmark graphs for testing community detection algorithms". In: *Physical review E* 78.4, p. 046110.

Latour, B. et al. (2012). "'The whole is always smaller than its parts'–a digital test of Gabriel Tardes' monads". In: *The British journal of sociology* 63.4, pp. 590–615.

Lawrence, R. Z. (1993). "Japan's different trade regime: an analysis with particular reference to Keiretsu". In: *The Journal of Economic Perspectives* 7.3, pp. 3–19.

Lazer, D. et al. (2009). "Life in the network: the coming age of computational social science". In: *Science* 323.5915, pp. 721–723.

Leydesdorff, L. (2007). "Betweenness centrality as an indicator of the interdisciplinarity of scientific journals". In: *Journal of the American Society for Information Science and Technology* 58.9, pp. 1303–1319.

Lincoln, J. R. and M. L. Gerlach (2004). *Japan's network economy: Structure, persistence, and change*. Vol. 24. Cambridge University Press.

Lincoln, J. R., M. L. Gerlach, and C. L. Ahmadjian (1996). "Keiretsu networks and corporate performance in Japan". In: *American sociological review*, pp. 67–88.

Lincoln, J. R., M. L. Gerlach, and P. Takahashi (1992). "Keiretsu networks in the Japanese economy: A dyad analysis of intercorporate ties". In: *American sociological review*, pp. 561–585.

Lincoln, J. R. and M. Shimotani (2009). "Whither the Keiretsu, Japan's Business Networks? How Were They Structured? What Did They Do? Why Are They Gone?" In: *Institute for Research on Labor and Employment*.

Louf, R., P. Jensen, and M Barthelemy (2013). "Emergence of hierarchy in cost-driven growth of spatial networks". In: *Proceedings of the National Academy of Sciences* 110.22, pp. 8824–8829.

McGuire, J. and S. Dow (2009). "Japanese keiretsu: Past, present, future". In: *Asia Pacific journal of management* 26.2, pp. 333–351.

Miwa, Y. and J. M. Ramseyer (2002). "The fable of the keiretsu". In: *Journal of Economics & Management Strategy*, pp. 169–224.

— (2006). *The fable of the keiretsu: urban legends of the Japanese economy*. University of Chicago Press.

Mucha, P. J. et al. (2010). "Community structure in time-dependent, multiscale, and multiplex networks". In: *Science* 328.5980, pp. 876–878.

Nakamura, M. (2002). "Mixed ownership of industrial firms in Japan: debt financing, banks and vertical keiretsu groups". In: *Economic Systems* 26.3, pp. 231–247.

Nakao, K. (1990). "Distribution of measures of centrality: enumerated distributions of Freeman's graph centrality measures". In: *Connections* 13.3, pp. 10–22.

Nepusz, T. et al. (2008). "Fuzzy communities and the concept of bridgeness in complex networks". In: *Physical Review E* 77.1, p. 016107.

Newman, M. E. J. (2003). "The structure and function of complex networks". In: *SIAM review* 45.2, pp. 167–256.

— (2005). "A measure of betweenness centrality based on random walks". In: *Social networks* 27.1, pp. 39–54.

— (2010). *Networks: an introduction*. Oxford University Press.

Padgett, J. F. and W. W. Powell (2012). *The emergence of organizations and markets*. Princeton University Press.

Palla, G., A.-L. Barabási, and T. Vicsek (2007). "Quantifying social group evolution". In: *Nature* 446.7136, pp. 664–667.

Peel, L. and A. Clauset (2015). "Detecting Change Points in the Large-Scale Structure of Evolving Networks." In: *AAAI*, pp. 2914–2920.

Peixoto, T. P. and M. Rosvall (2016). "Modeling sequences and temporal networks with dynamic community structures". In: *preprint arXiv:1509.04740*.

Rafols, I. (2014). "Knowledge Integration and Diffusion: Measures and Mapping of Diversity and Coherence". In: *Measuring Scholarly Impact*. Ed. by Y. Ding, R. Rousseau, and D. Wolfram. Springer, pp. 169–190.

Ramseyer, J. M. and Y. Miwa (2007). "The Good Occupation". In: *Harvard Law School Faculty Scholarship Series*, p. 2.

Rao, C. R. (1982). "Diversity and dissimilarity coefficients: a unified approach". In: *Theoretical Population Biology* 21.1, pp. 24–43.

Rosvall, M. and C. T. Bergstrom (2010). "Mapping change in large networks". In: *PloS one* 5.1, e8694.

Salton, G. and M. J. McGill (1983). *Introduction to modern information retrieval*. McGraw-Hill, Inc.

Scott, J. (1999). *Social network analysis*. Sage.

Shi, W. S. et al. (2014). "Domestic alliance network to attract foreign partners: Evidence from international joint ventures in China". In: *Journal of International Business Studies* 45.3, pp. 338–362.

Stirling, A. (2007). "A general framework for analysing diversity in science, technology and society". In: *Journal of the Royal Society Interface* 4.15, pp. 707–719.

Suehiro, A. (1992). "Capitalist development in postwar Thailand: Commercial bankers, industrial elite, and agribusiness groups". In: *Southeast Asian Capitalists*. Ed. by R. T. McVey. Ithaca, USA: Cornell University Press, pp. 35–63.

Tate, M. W. and L. A. Hyer (1973). "Inaccuracy of the X^2 test of goodness of fit when expected frequencies are small". In: *Journal of the American Statistical Association* 68.344, pp. 836–841.

Thomson Reuters. *Web of Science*. Restricted Access. URL: http://apps.webofknowledge.com/.

Toyo Keizai Shinposha (2000). *Kigyo keiretsu Soran [Survey of industrial conglomerates]*. (in Japanese). Tokyo.

— (2006). *Kaigai Shinshutsu Kigyo Soran [Survey of overseas subsidiaries]*. (in Japanese). Tokyo.

Venturini, T. and B. Latour (2010). "The social fabric: Digital traces and quali-quantitative methods". In: *Proceedings of Future En Seine 2009*, pp. 87–101.

Vespignani, A. (2009). "Predicting the behavior of techno-social systems". In: *Science* 325.5939, p. 425.

Weinstein, D. E. and Y. Yafeh (1995). "Japan's corporate groups: Collusive or competitive? An empirical investigation of keiretsu behavior". In: *The Journal of Industrial Economics*, pp. 359–376.

Xu, K. S., M. Kliger, and A. O. Hero III (2014). "Adaptive evolutionary clustering". In: *Data Mining and Knowledge Discovery* 28.2, pp. 304–336.

Yamashita, N., T. Matsuura, and K. Nakajima (2014). "Agglomeration effects of inter-firm backward and forward linkages: Evidence from Japanese manufacturing investment in China". In: *Journal of the Japanese and International Economies* 34, pp. 24–41.

Zhang, M. M. (2015). "Cross-national distance and insidership within networks: Japanese MNCs' ownership strategies in their overseas subsidiaries". In: *Asia Pacific Business Review* 21.3, pp. 403–423.