SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

HABILITATION A DIRIGER DES RECHERCHES DEMANDE D'AUTORISATION DE PRESENTATION EN SOUTENANCE

(A remplir par le candidat et à transmettre à la Commission des Habilitations qui se chargera de le transmettre au bureau des HDR)

Nom patronymique ROTH **Prénom**Camille

Nom marital

Né(e) le 23/07/1979 **à** Maisons-Laffitte (Yvelines)

Adresse 10 rue Demarquay, 75010 PARIS

Tél. 06 8034 0870

E-mail camille.roth@cnrs.fr

Situation : organisme employeur CNRS

Adresse 3 rue Michel-Ange, 75016 PARIS

Date de soutenance prévue 29 mai 2021

Titres Universitaires (préciser année et lieu d'obtention). Joindre photocopie du diplôme de doctorat.
Doctorat (2005, École Polytechnique)

Ingénieur des Ponts et Chaussées (2002, École des Ponts)

Sont proposés en qualité de <u>rapporteurs</u> :

NOM et Prénom	Qualité	Etablissement_Unité de recherche	Téléphone	Email
1 M FLEURY Éric	Professeur (informatique)	ENS Lyon Inria Paris	+33 672 162 974	eric.fleury@inria.fr
2 M GILBERT Nigel	Professeur (sociologie computationnelle)	University of Surrey (Royaume-Uni) Centre for Research in Social Simulation	+44 1483 689173	n.gilbert@surrey.ac.uk
3 M STROHMAIER Markus	Professor (informatique, chaire "Methods and Theories of Computational Social Sciences & Humanities")	RWTH Aachen (Allemagne) Department of Computer Science	+49 241 80 25478	markus.strohmaier@humtec.rwth-aachen.de
4 M NADAL Jean-Pierre	Directeur de recherche (physique)	CNRS CAMS (Centre d'Analyse et de Mathématique sociale)	+33 1 44 32 32 75	jean-pierre.nadal phys.ens.fr
5 M				
6 M				

Le candidat propose de 4 à 6 rapporteurs. Il ne doit pas avoir co-signé d'articles avec les rapporteurs dont il propose les noms, ni avoir collaboré avec eux. Les rapporteurs ne doivent pas appartenir à l'établissement de rattachement du candidat ni à son unité de recherche.

Quand la commission des HDR émet un avis favorable, elle désigne trois rapporteurs à qui elle demande leur rapport. La commission des HDR envoie le manuscrit aux trois rapporteurs accompagné du document de rapport. A la réception de l'ensemble des rapports, ils seront communiqués au candidat.

Sont proposés en qualité de membres du jury :

Le candidat adresse à la commission des HDR une proposition de jury qui est constitué d'au moins 5 membres HDR ou, le cas échéant, de personnalités françaises ou étrangères retenues en raison de leurs compétences scientifiques dont au moins 2 sont rapporteurs du mémoire d'habilitation. La moitié au moins des membres du jury est composée de professeurs ou assimilés. La moitié au moins des membres doit être composée de personnalités françaises ou étrangères extérieures à l'établissement (cf. arrêté du 13 février 1992). Le jury doit également comprendre un membre de la communauté Sorbonne Université habilité à diriger des recherches.

NOM et Prénom	Qualité	Etablissement_Unité de recherche	Téléphone	Email
1 M BERESTYCKI Henri	Directeur d'études (mathématiques)	EHESS CAMS (CNRS/EHESS)	+33 1 49 54 20 33	hb@ehess.fr
2 M FLEURY Éric	Professeur (informatique)	ENS Lyon Inria Paris	+33 672 162 974	eric.fleury@inria.fr
3 M GILBERT Nigel	Professeur (sociologie computationnelle)	CNRS LIP6	+44 1483 689173	n.gilbert@surrey.ac.uk
4 M LAZEGA Emmanuel	Professeur (sociologie)	Sciences Po Paris Centre de Sociologie des organisations	+33 1 40 62 65 70	emmanuel.lazega@sciencespo.fr
5 Mme MAGNIEN Clémence	Directrice de recherche (informatique)	CNRS LIP6 (Sorbonne Université / CNRS)	+33 1 44 27 87 92	clemence.magnien@lip6.fr
6 Mme SAINT-CHARLES Johanne	Professeure (communication)	Université du Québec à Montréal Département de communication sociale et publique	+1 514 987-3000 ext. 2081	saint-charles.johanne@uqam.ca
7 M STROHMAIER Markus	Professeur (informatique)	RWTH Aachen (Allemagne) Department of Computer Science	+49 241 80 25478	markus.strohmaier@gesis.org
8 M STUMME Gerd	Professeur (informatique)	Universität Kassel (Allemagne) Knowledge and Data Engineering	+49 561 804-6251	stumme@cs.uni-kassel.de

Le Candidat

Camille ROTH

Date et signature

08 avril 2021

C. Roth

Pour le Président,
la Commission des Habilitations à diriger des recherches

Date, tampon et signature

12/04/2021

SORBONNE UNIVERSITÉ
Commission des Habilitations à Diriger les Recherches (HDR) de l'UFR d'Ingénierie

Mme Hélène ROUSSEL

Sorbonne Université
Bureau des HDR
Esc. G, 2ème étage
15 rue de l'école de médecine
75270-PARIS CEDEX 06

Tél. pour les candidat de A à EL : 01 44 27 28 05
Tél. pour les candidats de EL à MON : 01 44 27 28 07
Tél. pour les candidats de MOO à Z : 44 27 28 03
E-mail : scolarite.doctorat@sorbonne-universite.fr

Socio-Semantic Systems

Camille ROTH

Habilitation à Diriger des Recherches

Spécialité: Informatique

SORBONNE UNIVERSITÉ

Jury

M. Henri BERESTYCKI	examinateur	Directeur d'études	mathématiques	EHESS
M. Éric FLEURY	rapporteur	Professeur	informatique	ENS Lyon / Inria Paris
M. Nigel GILBERT	rapporteur	Professeur	sociologie computationnelle	Surrey (Royaume-Uni)
M. Emmanuel LAZEGA	examinateur	Professeur	sociologie	Sciences Po Paris
Mme Clémence MAGNIEN	examinatrice	Directrice de recherche	informatique	CNRS
Mme Johanne SAINT-CHARLES	examinatrice	Professeure	communication	UQÀM (Canada)
M. Markus STROHMAIER	rapporteur	Professeur	informatique	RWTH Aachen (Allemagne)

Abstract

Socio-semantic systems involve actors who create and process knowledge, exchange information and create ties between ideas in a distributed and networked manner: we-bloggers and micro-bloggers, communities of scientists, of software developers and of wiki contributors are, among others, examples of such systems. The state of the art in this regard focuses on two main issues which are increasingly addressed in an interdependent manner: the description of content dynamics and the study of interaction network characteristics and evolution. The present document contextualizes and reviews my efforts, over more than a decade, at merging both types of dynamics into co-evolutionary, multi-level modeling frameworks, where social and semantic aspects are being jointly appraised by using socio-semantic graphs, hypergraphs and lattices. It also evokes related streams of research dealing with more specific topics, such as diffusion processes and authority phenomena, and theoretical issues, such as the modeling of graph structure and dynamics in a variety of empirical and formal contexts.

Acknowledgements

My research program would assuredly not have been possible without all the interactions I had with many inspiring colleagues over the last two decades. I would particularly like to thank key co-authors, team members including Telmo Menezes and Jean-Philippe Cointet, as well as Chih-Chun Chen, Floriana Gargiulo, Sébastien Lerique, Bivas Mitra, Jérémie Poiroux, Lionel Tabourier, and Carla Taramasco; project partners Frédéric Amblard, Marc Barthélémy, Nikita Basov, Dominique Cardon, David Chavalarias, Guilhem Fouetillou, Klaus Hamberger, Matthieu Latapy, Clémence Magnien, Sergei Obiedkov, Christophe Prieur, and Dario Taraborelli; and, last but not least, the people I had the pleasure to be under the supervision of early in my career: Paul Bourgine, David A. Lane and Nigel Gilbert.

Contents

Foreword

The core of my research program has focused on the question of *social cognition*. This term should be understood as the socially distributed processing of knowledge within a sizable system of interacting individuals. In this regard, it is less connected to the traditional acceptation of the term in psychology which focuses rather on the cognitive processes involved at the strict level of human interactions. This question has principally been a social science concern for much of the 20th century, especially in cultural anthropology and sociology. The last two decades have witnessed an increasing use of formal methods borrowing to computer science, applied mathematics, statistical physics, to the end of complex system modeling and large dataset analysis. This convergence is currently widely labeled as "computational social science" and quite often focuses on large socio-technical systems where actors interact in a relatively decentralized and autonomous manner and, to some extent, rely on information and communication technologies which eventually host, organize and facilitate large-scale social cognition.

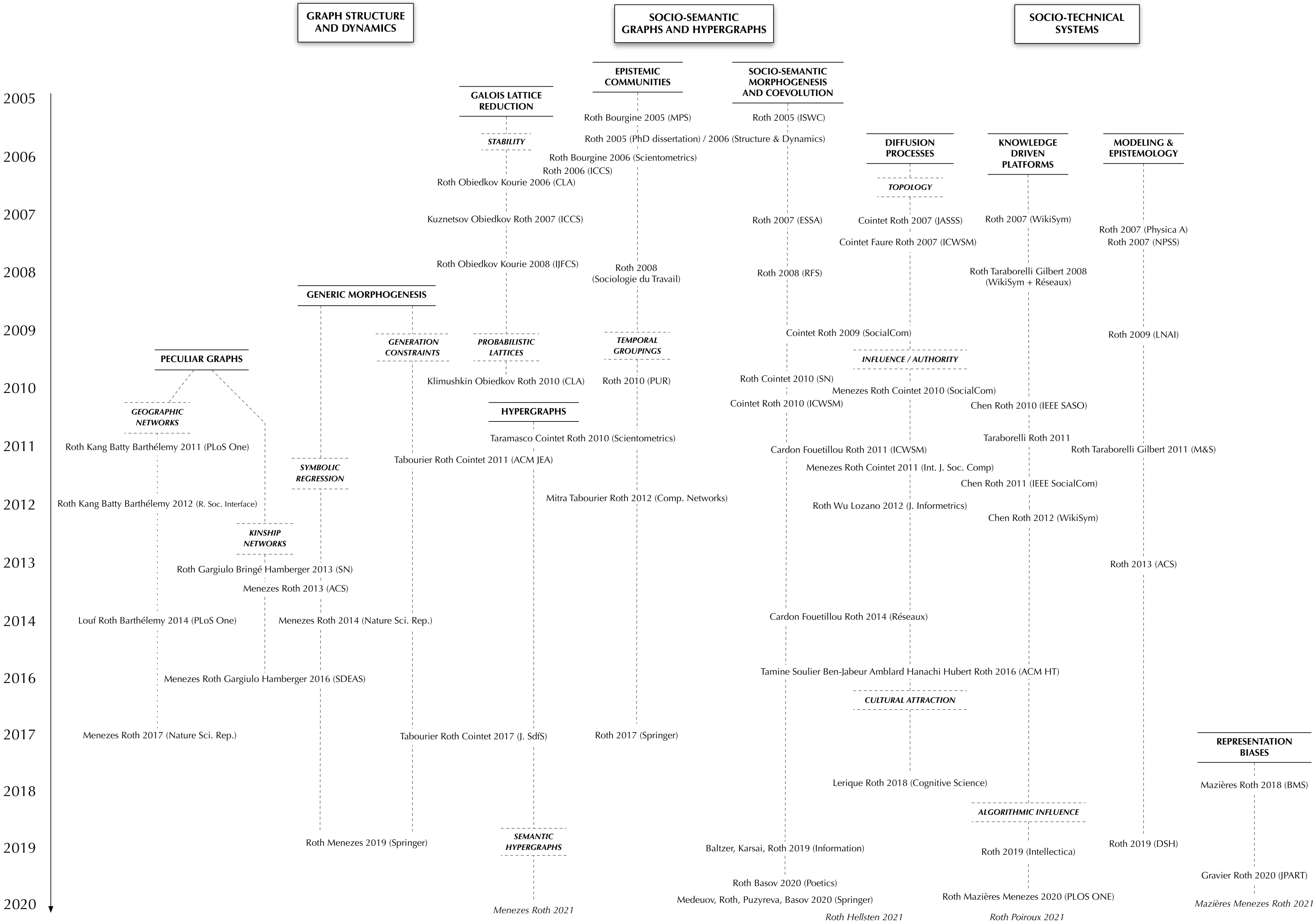
Science and its various subcommunities long provided the only sizable system whose dynamics could be quantitatively appraised as such, and to which a whole discipline has been almost exclusively devoted: scientometrics, as a precursor of big (social) data. While this system has mainly relied on a seemingly simple socio-technical apparatus, revolving around cultural artifacts called "books", synchronized repositories called "libraries" and physical gatherings called "conferences", data and results have always been collected and processed in a mainly asynchronous and distributed manner, by scientific teams operating locally, on subproblems, with no central plan in mind. This monopoly as a perfect playground for the empirical study of social cognition progressively ceased as a myriad of online user-generated content platforms flourished — including blogs, wikis, open-source development platforms, social networking sites, tagging platforms. The advent of these socio-technical systems not only facilitated social cognition processes per se, they made their systematic in vivo observation much easier as well.

As a result, it comes to no surprise that most of my empirical interests revolved around science and online communities. Notwithstanding and in all generality, two elements of these systems, which I shall from now on denote as *socio-semantic systems*, are essential to social cognition processes: on one side, agents interacting in diverse manners and, on the other side, a mesh of informational "items" or knowledge artifacts (texts, opinions, tags, and more broadly digital content). By the turn of the 2000s, much of cultural sociology was already strongly aware of this duality, yet most the quantitative literature on real-world social cognition relied on formal frameworks that primarily focused on one realm (the social or the semantic structure). Empirical scholarship that attempted to bind the two realms was relatively limited and often restricted to bipartite relationships.

This manuscript principally aims at framing and reviewing my contributions, over more than one decade, to the description and modeling of socio-semantic systems. It concomitantly advocates the notion that the full appraisal of social cognition phenomena requires modeling frameworks which jointly feature social structure and semantic characteristics; that is, socio-semantic frameworks, and, above all, *socio-semantic hypergraphs*. Furthermore, a specific focus of my research relates to the co-evolution between the social and semantic dimensions, as interaction and information dynamics often appear to obey similar time-scales, especially in ICT-based socio-technical systems: virtually by design, content production indeed involves interactions which contribute to shape future content creation which, in turn, influences the evolution of the social fabric; and so on.

This memoir begins with a contextualization of the main relevant research streams in Section 1, from social cognition (1.1) to cultural sociology (1.2), up to computational social science (1.3). I shall then focus on areas where I have been particularly active and which variously contribute to the operational study of social cognition. In Section 2, I address the broader issue of network morphogenesis, looking for rules and constraints likely to explain the structure and dynamics of a given interaction system, considering either social networks (2.1) or relatively atypical social science networks (namely, geographical networks and kinship networks, 2.2). Section 3 further deals with socio-semantic modeling frameworks, principally structured around the introduction of the notion of socio-semantic networks of various types: socio-semantic graphs, hypergraphs and lattices. Lastly, Section 4 discusses more specific phenomena such as information propagation, socio-semantic authority and position analysis, referencing and evaluation behavior, which occur within or on top of a series of socio-technical systems, mostly in a digital or scientific context.

The diagram on the next page summarizes the temporal articulation of my peer-reviewed contributions (see the bibliography for precise references).



GRAPH STRUCTURE AND DYNAMICS
SOCIO-SEMANTIC GRAPHS AND HYPERGRAPHS
SOCIO-TECHNICAL SYSTEMS
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2016
2017
2018
2019
2020
GALOIS LATTICE REDUCTION
EPISTEMIC COMMUNITIES
SOCIO-SEMANTIC MORPHOGENESIS AND COEVOLUTION
Roth Bourgine 2005 (MPS)
Roth 2005 (ISWC)
STABILITY
Roth 2005 (PhD dissertation) / 2006 (Structure & Dynamics)
DIFFUSION PROCESSES
KNOWLEDGE DRIVEN PLATFORMS
MODELING & EPISTEMOLOGY
Roth Bourgine 2006 (Scientometrics)
Roth 2006 (ICCS)
Roth Obiedkov Kourie 2006 (CLA)
TOPOLOGY
Kuznetsov Obiedkov Roth 2007 (ICCS)
Roth 2007 (ESSA)
Cointet Roth 2007 (JASSS)
Roth 2007 (WikiSym)
Roth 2007 (Physica A)
Roth 2007 (NPSS)
Cointet Faure Roth 2007 (ICWSM)
Roth Obiedkov Kourie 2008 (IJFCS)
Roth 2008 (Sociologie du Travail)
Roth 2008 (RFS)
Roth Taraborelli Gilbert 2008 (WikiSym + Réseaux)
GENERIC MORPHOGENESIS
GENERATION CONSTRAINTS
PROBABILISTIC LATTICES
TEMPORAL GROUPINGS
Cointet Roth 2009 (SocialCom)
Roth 2009 (LNAI)
PECULIAR GRAPHS
INFLUENCE / AUTHORITY
Klimushkin Obiedkov Roth 2010 (CLA)
Roth 2010 (PUR)
Roth Cointet 2010 (SN)
Menezes Roth Cointet 2010 (SocialCom)
GEOGRAPHIC NETWORKS
HYPERGRAPHS
Cointet Roth 2010 (ICWSM)
Chen Roth 2010 (IEEE SASO)
Taramasco Cointet Roth 2010 (Scientometrics)
Taraborelli Roth 2011
Roth Kang Batty Barthélemy 2011 (PLoS One)
Cardon Fouetillou Roth 2011 (ICWSM)
Roth Taraborelli Gilbert 2011 (M&S)
Tabourier Roth Cointet 2011 (ACM JEA)
SYMBOLIC REGRESSION
Menezes Roth Cointet 2011 (Int. J. Soc. Comp)
Chen Roth 2011 (IEEE SocialCom)
Mitra Tabourier Roth 2012 (Comp. Networks)
Roth Kang Batty Barthélemy 2012 (R. Soc. Interface)
Roth Wu Lozano 2012 (J. Informetrics)
Chen Roth 2012 (WikiSym)
KINSHIP NETWORKS
Roth 2013 (ACS)
Roth Gargiulo Bringé Hamberger 2013 (SN)
Menezes Roth 2013 (ACS)
Cardon Fouetillou Roth 2014 (Réseaux)
Louf Roth Barthélemy 2014 (PLoS One)
Menezes Roth 2014 (Nature Sci. Rep.)
Tamine Soulier Ben-Jabeur Amblard Hanachi Hubert Roth 2016 (ACM HT)
Menezes Roth Gargiulo Hamberger 2016 (SDEAS)
CULTURAL ATTRACTION
Menezes Roth 2017 (Nature Sci. Rep.)
Tabourier Roth Cointet 2017 (J. SdfS)
Roth 2017 (Springer)
REPRESENTATION BIASES
Lerique Roth 2018 (Cognitive Science)
Mazières Roth 2018 (BMS)
ALGORITHMIC INFLUENCE
Roth Menezes 2019 (Springer)
SEMANTIC HYPERGRAPHS
Baltzer, Karsai, Roth 2019 (Information)
Roth 2019 (Intellectica)
Roth 2019 (DSH)
Gravier Roth 2020 (JPART)
Roth Basov 2020 (Poetics)
Roth Mazières Menezes 2020 (PLOS ONE)
Medeuov, Roth, Puzyreva, Basov 2020 (Springer)
Menezes Roth 2021
Roth Hellsten 2021
Roth Poiroux 2021
Mazières Menezes Roth 2021

1 Articulating structure and culture

It would certainly be way beyond the scope of the present manuscript to review all the possible scientific areas that relate in some way to social cognition — a good part of social science and humanities has actually to deal, at least implicitly, with social cognitive processes. I shall rather focus on the three main fields that provide an acutely pertinent epistemological background to their formal and quantitative understanding: social epistemology, cultural anthropology, and cultural sociology. This section will then close on the empirical efforts observed concurrently in fields related to natural and formal sciences, first and foremost dealing with social complex system modeling and computational social science.

1.1 Social cognition

Two fields are directly concerned with the articulation of cognitive processes per se and the distribution production of knowledge, bordering on analytical philosophy and anthropology.

Social epistemology is perhaps the only area of research to explicitly and almost exclusively address the conditions of the collective *foundation* of knowledge. Its most theoretical ramifications deal with the characterization of what may be collectively accepted as knowledge, by defining for instance a given proposition p as "community knowledge" *iff* agents know p and know that others know p and trust them (Kitcher, 1995) and typically overlap with epistemic logic (which involves, however, little empirical modeling – see nevertheless Shoham and Leyton-Brown (2008), for instance, for a normative application to agent-based models).

The more sociological ramifications focus on the social factors behind the construction and adoption of knowledge, regarding for example the origin of consensual or authoritative statements (Lazega, 1992). These works basically question the influence of the bias induced by actors, voluntarily or not, on the processing of information and its evaluation by a given social group. Here, science appears to be a natural prototype (Knorr-Cetina, 1999), with early sociological studies dealing with the joint dynamics of knowledge and social organization (Latour and Woolgar, 1979). This sociological stance leads in particular to the study of the social procedures pertaining to the organization of cognitive labor. This is more closely connected to socio-technical systems because of the specific emphasis on the role of the technological environment: Hutchins (1995), for one, exemplified the notion of "distributed cognition" by showing that the successful piloting of a ship to seaport requires a distributed effort where all parties, humans and devices alike, have to play a local role — science also represents a typical case, again, and is interestingly illustrated in the so-called "actor-network theory" ontology (Callon et al., 1986), where scientific agents and artifacts are indistinctly gathered into a hybrid network.

Cultural anthropology addresses the conditions surrounding the propagation and reproduction of knowledge and representations: it shares similar high-level goals with social epistemology, yet with a specific focus on the emergence of culture and cultural similarity: for instance, "explaining the capacity of some representations to propagate until becoming precisely cultural, that is, revealing the reasons of their contagiosity" (Lenclud, 1998). The articulation with social cognition appears even more evidently when culture is defined as "acquired information, such as knowledge, beliefs, and values, that is inherited through social learning, and expressed in behaviors and artifacts" (Mesoudi et al., 2004). This research stream crucially focuses on social learning and how knowledge and practices are transmitted until they become widespread (and thus, cultural). One of the foundational puzzles relates to the tension between the observed poor quality of inter-individual transmission (understanding and reformulating what is being heard or seen is indeed typically a quite noisy process) and the remarkable macro-level observation of coherence and cultural clusters i.e., wide arrays of actors sharing the same beliefs or social behaviors.

This research program is related to several notable proposals enabling a modeling approach. Memetics (Dawkins, 1976) in particular has long been seen by modelers as an efficient and naturalistic framework for understanding cultural convergence (see for instance Conte, 2000), even though it raised doubts from the side of cultural anthropology itself, concerning in particular the assumption that there exist atomic (cultural) representations and high-fidelity replication. The theory of cultural epidemiology or cultural contagion proposed by Sperber (1996) received a wider anthropological support, by clarifying the underlying cognitive processes and, notably, emphasizing both the role of reformulation and the existence of aggregates of cultural representations rather than cultural "atoms" (thereby extending the Levi-Straussian notion that a "myth" is "the set of all of its versions"). In any case, here again, modeling efforts, although convincing, have generally remained less descriptive than normative (see e.g., Claidière and Sperber, 2007). They are also, and perhaps more importantly, a specific formulation of the issue of social cognition in two respects: by aiming principally at explaining the emergence of culture as such, and by not giving much weight to the (heterogeneous) structure of social interactions in the dynamics of content and representations.

1.2 Cultural sociology

On the whole, while the two above fields provide conceptual guidance on cultural dynamics in its most theoretical and qualitative aspects, they seem yet to have yielded a relatively limited literature in terms of empirical models. Cultural sociology appears to have gone farther in this direction. For several decades this field has insisted on the duality of structure and culture, under the impulse of a number of social scientists generally concerned with social network analysis (SNA). Already in the 1990s, calls to jointly study both realms could build upon a wealth of pioneering intuitions and works revolving around three main ideas: the critique of a strict social structuralism, the dependence of networks on meaning, and the need to introduce the cultural level directly into SNA.

Sewell (1992) recalls how sociology and, more precisely, a large part of SNA typically diverge from anthropology (or more broadly "semiotically inclined" social science) in the interpretation of what counts primarily as "structure". For the former, the "material" (or interactional) realm prevails over the "mental" (or cultural) realm, which is thus secondary. The opposite holds for the latter. In effect, SNA has a long history of uncovering and describing categories and social order from the structure first, rather than using preexisting social or symbolic categories (see also Section 3.3.2), even if typical node patterns such as structural equivalence (Lorrain and White, 1971) have more to do with meaning than cohesion, and even though early social network analyses could mobilize cultural features in their depictions (Fuhse, 2009). Assuming the primacy of structure over culture or, more concretely, of relations over "categories" (in the broad sense) may lead to overlook or at least neglect the crucial contribution of the cultural and semiotic dimensions to the formation of stucture (Emirbayer and Goodwin, 1994). Put differently, cultural sociology forcefully rejects a strong structural viewpoint whereby networks alone would somehow reveal enough about the constraints that apply to individuals, while the study of motives, beliefs and identities should be left to other fields, be it anthropology or psychology (Vaisey and Lizardo, 2010). Pachucki and Breiger (2010) further insist that "culture prods, evokes, and constitutes social networks" and that the "structural presence –or absence– of ties may have cultural explanations as well": a social network depends on identities, contexts, meanings. White (1992) even speaks of "networks of meanings" to the point that each identity, each interaction context configures a different network: "persons come into existence and are formed as overlaps among identities from distinct network-populations", which echoes the earlier notion of "multi-purpose ties" introduced by Fine and Kleinman (1983) and the later conflict-based actor networks of Tilly (1997). Godart and White (2010) provide a more specific explanation for this interconnection in the form of "netdoms", a term which stands for "social *net*works" and "semiotic *dom*ains". Netdoms crucially rely on the concept of "stories", construed as sets of semantic relationships which may precede actual interactions, as is for instance the case in kinship or organizational networks (e.g., "grand-grand-uncle" or military ranks). Albeit principally programmatic, their proposal provides an account of how meanings travel from netdoms to becoming institutions, and back; how stories, meaning structures, social positions, and social structures are intertwined and may coevolve diachronically. This systemic and potentially autopoietic perspective suggests that networks are not only influenced by meaning but are also based on meaning (Fuhse, 2009), and more provocatively that social networks and culture could even just be secondary and joint manifestations of a "deeper set of forces" (Breiger, 2010).

The recent review by Ferguson et al. (2017) demonstrates that calls to simultaneously address meaning and interaction structures remain very much relevant. Notwithstanding, operational endeavors to incorporate the cultural level into SNA also have a long tradition. These efforts roughly appear to follow the guidelines of the "strong program for cultural sociology" advocated by Alexander and Smith (2001), who envision structuralism and hermeneutics as separate yet cross-fertilizing approaches — in effect, an integrated dualism that acknowledges the intertwinement of structure and culture while keeping them "empirically and analytically distinct" (Vaisey and Lizardo, 2010). The seminal work of Breiger (1974) inaugurated the appraisal of this duality in SNA by introducing so-called "membership networks", which connect individuals to groups they are member of, whereby groups may indifferently represent concrete collectives (say, committees, organizations) or more abstract entities (say, events, identities). These networks rely therefore on a bipartite connection between purely social nodes (individuals) and more semantic nodes (affiliations) which may further be projected as pure social or semantic networks — where links figure co-membership, either of individuals or affiliations. At the meta-level, the bipartite connection between individuals and affiliations makes it possible to define a broader notion of groups by dually connecting the social and semantic levels in a hierarchical structure known as Galois Lattices (Freeman and White, 1993; Mohr and Duquenne, 1997). These lattices, which I will address in more detail in Section 3.4, jointly gather subsets of actors sharing the same subsets of properties (affiliations, domains of interest, participation in events). Not only do they constitute a formalization of structural equivalence in a bipartite context (individuals belong to the same group if they are connected in the same way to some properties, instead of the same actors), but they also prefigure a way of appraising inter-actor relationships stemming from similar patterns of symbolic affiliations, which is reminiscent of Bourdieu (1985)'s duality between social (power) relationships and symbolic (power) relationships.

In these approaches, inter-actor relationships are derived from joint affiliations, rather than directly observed from interactions. The social graph resulting from the projection of the bipartite actor-property graph onto actors exemplifies the fundamental dichotomy between interaction and affiliation, between monopartite and eminently bipartite networks. In this regard, the pioneering work of Carley (1986) introduces a framework that precisely comprises all the building blocks required to jointly appraise the social and the semantic structure: first by defining them separately, then by connecting them. Her work is settled in the context of collective decision-making: a social cognitive process where the analysis of facts by individuals heavily depends on social structure, locally and collectively. The social structure takes the typical form of a social network whose clusters are block-modeled (using the classical CONCOR approach of Breiger et al., 1975) and characterized in terms of intra- and inter-group density. The semantic structure is figured by concept maps i.e., semantic networks. These graphs are also subjected to the computation of typical SNA measures such as centrality and transitivity, albeit interpreted from a cognitive perspective. More importantly, actor-centric concept maps may be studied, compared, and intersected for pairs or groups of actors, which achieves the hybridation of connections of all types — social, semantic, and socio-semantic. Another type of successful operationalization consists in statistical modeling frameworks for the coevolution of structure and behavior (Leenders, 1997; Snijders et al., 2007), where the contribution of both behavioral and structural properties in the formation of new links is being estimated within a single model. This paved the way for a class of exponentional random graph models of multiplex and multilevel networks, which aim at characterizing hybrid patterns made of different types of links that bear different meanings. Multiplex networks rely on a single type of nodes (say, actors) but accommodate a variety of semantically-labeled social links (Lazega and Pattison, 1999). Multilevel networks, by contrast, rely on two types of nodes (actors at the micro-level, and say, groups, organizations, or concepts, at the macro-level): patterns are hybrid in that they mix distinct types of monopartite (actor-actor, group-group) or bipartite (actor-group) links. The non-actor realm may consist of organizations (Lazega et al., 2008; Wang et al., 2013) or, more recently, concepts (Basov and Brennecke, 2017).

1.3 Social complex systems and computational social science

A last relevant research strand construes socio-cognitive dynamics as a social complex system (Holland, 1996; Sawyer, 2001), focusing notably on the articulation between macro-level phenomena and micro-level processes. This literature participates more broadly to the understanding of human dynamics from a natural science perspective which, over the last two decades, has been variously affiliated with fields such as "social simulation" (Gilbert and Troitzsch, 1999), "social computing" (Wang et al., 2007) and, increasingly, "computational social science" (CSS, see Lazer et al., 2009; Conte et al., 2012).

It has now become commonplace to say that these approaches experienced an unprecedented growth in a large part as a result of the ubiquitous availability of sizable datasets (fashionably called "big data") detailing the in vivo traces of human behavior — especially thanks to digital devices which further affect social science methods as a whole (Ruppert et al., 2013). Worth noting, nonetheless, is that much of this literature generally consists of "data science" in the very sense that data is, indeed, "data" (latine for "given"), rather than "fata" (made): empirically, computational social scientists oftentimes have little control over the production process of their data. It thus overlaps with issues related to "social sensing", by considering actors as a distributed network of sensors enabling the knowledge (Ginsberg et al., 2009) or prediction (Asur and Huberman, 2010) of the state of a given social system and, more broadly, by offering the opportunity to reverse-engineer human behavior from its traces. These efforts relate to a larger body of quantitative studies dealing on one hand with social interactions at large, and on the other hand with content dynamics analysis.

Characterizing human networks. As suggested in the previous section, the analysis of social networks has already had a long history made of mathematical sociology studies, starting essentially from the 1940s. The period until the late 1990s was nevertheless rather focused on "small" case-studies, with datasets describing social interactions of groups of less than a hundred people, more often a few dozens. It offered the opportunity to introduce most of the key formal frameworks and measures of today: centralities, comparisons with random networks, behavioral inference, community detection — the classic book of Wasserman and Faust (1994) reviews the already rich state of the art of this field as of the mid-1990s. The more recent years have witnessed the emergence of large-scale studies of human-related networks, in the broad sense, as part of a broader effort for the study of so-called "complex networks", stemming principally from statistical physics and computer science. This dynamics had been initially fueled by the recurrent observation that empirical networks were rather heterogeneous (with keywords such as "scale-free", "power-law", etc.) and cohesive (with keywords such as "small worlds", "clustering" or "clusters"), and partly by the quest for universal laws applying to all types of empirical interaction structures. In passing, this offered the opportunity to revisit some of the earlier mathematical sociology concepts on much larger as well as much more diverse datasets — see e.g., the very early review by Albert and Barabási (2002), and more targeted reviews on a variety of subsequent issues pertaining to, inter alia, communities (Fortunato, 2010), spatial networks (Barthélemy, 2011), and more recently graph embeddings (Goyal and Ferrara, 2018) or non-dyadic interaction structures such as hypergraphs (Battiston et al., 2020). On the whole, it is reasonable to claim that these overlapping and, now, merging streams of research —SNA, complex system science and CSS— have achieved today a satisfying characterization of empirical social networks, both statically and dynamically. After the initial all-purpose, universal models targeting the reconstruction of ubiquitous heterogeneous degree distributions observed in almost all systems, realistic morphogenesis models have successfully been proposed in increasingly specific case studies in order to explain increasingly specific patterns. Section 2.1 will offer a detailed account of the state of the art on social network morphogenesis.

Characterizing semantic configurations. Studies on science and more specifically scientometrics were among the first research areas to develop a massive effort of quantitative mapping of the distribution of knowledge in social systems. Using large bibliographic corpora, they proposed history reconstruction methods based on clusters of terms or authors, principally by examing co-citation (McCain, 1986), co-word/co-occurrence (Callon et al., 1991) or collaboration maps (Börner et al., 2003) that ultimately aimed at describing the relative arrangement of scientific subfields in terms of their main topics automatically extracted from textual data.

In current CSS research, the quantitative analysis of natural language is referred to by a variety of umbrella terms such as "text mining" (Srivastava and Sahami, 2009; Evans and Aceves, 2016), "automated text analysis" (Grimmer and Stewart, 2013) or "text-as-data methods" (Wilkerson and Casas, 2017), which exhibit a wide range of sophistication, from simple numerical statistics to more elaborate machine learning algorithms. Some methods essentially produce scalar numbers, for instance to measure text similarity (e.g., the traditional cosine distance, see Singhal, 2001) or valence, as in sentiment analysis (Pang et al., 2008) which evaluates the positive or negative emotional content of a text (e.g., public sentiment towards political candidates in social media, Wang et al., 2012). Texts may also be projected on a fixed number of categories, as in the case of ideological estimation (e.g., political leaning in terms of left or right, see Sim et al., 2013). Other approaches preserve the level of words and n-grams. Some aim at the extraction of salient terms, using helper measures such as the famous "tf.idf" product (term frequency times inverse document frequency, Salton and Buckley, 1988) or co-occurrence centrality (TextRank, Mihalcea and Tarau, 2004). Others deal with the recognition of so-called "Named Entities" (i.e., people, locations, organizations; see Nadeau and Sekine, 2007), for instance from news corpora (Diesner and Carley, 2005) or Twitter streams (Ritter et al., 2011) or with the extraction of chunks of text matching certain patterns, for example p-values from scientific articles (Chavalarias et al., 2016). This level enables applications akin to signal analysis, aiming for instance at the description of dynamical classes of word use, in terms of spikes (Gruhl et al., 2004), heterogeneity (Cattuto et al., 2007) or similar temporal (Lehmann et al., 2012) and competitive (Weng et al., 2012) dynamics.

Another strand of techniques operates at the level of word sets, such as the above-mentioned family of co-word maps (for longitudinal, multilevel examples, see Cui et al., 2011; Shahaf et al., 2013), as well as so-called distributional topic models. This includes the very popular Latent Dirichlet Allocation (LDA, Blei, 2012) which characterizes topics by their probability of generating each of the words found in a set of documents, which are thus seen as a random mixture of topics. LDA uses a generative process to statistically infer these probabilities: topics are ultimately described by a (weighted) set of words, for instance "EU, Ursula von der Leyen, Boris Johnson, Barnier, Trade", which human observers can use to infer some higher-level topic label, such as "Brexit Negotiations". A more recent approach consists in applying stochastic blockmodeling, which originally stems from pure SNA, in order to discover joint groups of documents which use keywords in a similar fashion (Gerlach et al., 2018). The jury is still out as to whether distributional approaches are more efficient than co-occurrence and co-word maps (Leydesdorff and Nerghes, 2017). Recent advances in word embedding techniques (Mikolov et al., 2013) have also made it possible to describe topics extensionally as clusters of documents in some properly defined space (Le and Mikolov, 2014; Angelov, 2020).

Finally, a rather recent literature focuses on the sentence level, beyond word and n-gram distributions. Early on, Leskovec et al. (2009) described clusters of relatively similar quotations of public figures reformulated by bloggers with some amount of variation and reformulation. This enables the description of low-level processes of sentence evolution (Simmons et al., 2011), which in passing fulfills some of the preliminary objectives of cultural epidemiologists. There is currently an increasing trend of dealing with the syntactic structure to abstract meaning from sentences. This includes the field of relationship extraction from free text (Van Atteveldt et al., 2008; Mausam et al., 2012; Angeli et al., 2015; Vossen et al., 2016) and argumentation mining (Lippi and Torroni, 2016) which makes extensive use of machine learning to extract claims, arguments and premises. For instance, Ruiz et al. (2016) propose a system aimed at analyzing claims in the context of climate negotiations. It leverages dependency parse trees and general ontologies to extract tuples of the form $\langle$actor, predicate, negotiation point$\rangle$ where actors are stakeholders (e.g., countries), predicates express agreement or opposition, and negotiation points are identified by chunks of text. Similarly, Van Atteveldt et al. (2017) extract source-subject-predicate clauses from news reports to show differences in framing patterns between distinct sources.

Intertwining both. As seen in the previous subsection, mathematical and cultural sociology had already posed, at the end of 20th century, the issue of a co-evolution, or at least a correlation, between social and semantic, or behavioral aspects. On the complex system modeling side, attempts are more recent and started with the characterization of what may be called "unidirectional" correlations: more precisely, by showing how information distribution may explain or shed light on the social structure, which is appraised as a dependent variable of the semantic structure. This concerned especially two main questions: fragmentation and selection; and correspondingly yielded one main type of insights: coloring either clusters or dyadic connectedness with the help of semantic features and semantic similarity. This literature typically refers to the seminal sociological work of McPherson et al. (2001) on homophilic behavior.

At the macro-level, social network clusters have been shown to be generally semantically homogeneous, in that they exhibit similar political leanings (Adamic and Glance, 2005; Himelboim et al., 2013), geographical properties (Etling et al., 2010) or both (Hoffmann, 2014); even though the strength of this observation may also heavily depend on link types (Conover et al., 2011), linguistic features (Kim et al., 2014; Eleta and Golbeck, 2014), topics (Barberá et al., 2015; Garimella et al., 2016; Himelboim et al., 2017), and time (Gaumont et al., 2018).

At the micro-level of dyadic edges, the formation of citation (McGlohon et al., 2007; Cointet and Roth, 2009), interaction (Schifanella et al., 2010; Kossinets and Watts, 2009) or affiliation (Backstrom et al., 2006) links also depend on semantic features, generally exhibiting homophily. Here again, there are differences across distinct network types and topics (Lietz et al., 2014; Tyagi et al., 2020; Cinelli et al., 2021), whereby interaction is generally less homophilic than affiliation or citation. Similarly, in a variety of networks, semantic similarity depends on the strength of structural connectedness, both statically (Mitzlaff et al., 2013) and diachronically (Crandall et al., 2008).

Dynamic aspects have been increasingly addressed by this literature and directly appeal to the co-evolutionary nature of structure and culture. Of particular interest are the questions of the dependence of information and content adoption on structure, and of the generation and configuration of socio-semantic clusters, both in a normative and in an empirical setting. I shall review these issues in much more detail in Sections 3 and 4.

2 Graph morphogenesis

An overarching issue concerns the study of graph structure and dynamics in all generality, as a precondition to the understanding of the role of networks in appraising specific case studies. Before delving deeper into the topics evoked in the previous chapter and addressing the socio-semantic case in the next chapter, this chapter thus aims at reviewing the general issue of social graph morphogenesis, both in the most typical situation, where networks are both sparse and heterogeneous (2.1), and in the rather non-conventional case of weighted, dense networks occurring in peculiar contexts such as mobility and kinship networks (2.2).

2.1 Generic case: sparse and heterogeneous networks

The modeling of network morphogenesis has generated a particularly substantial literature following a series of findings in the early 2000s where most real-world networks appeared to exhibit universal topological features: high level of clustering, existence of modules or clusters, and heterogeneous connectivity distribution (roughly following power or log-normal laws). The corresponding state of the art may essentially be organized according to two key dichotomies: the first one relates to the *target* of models, the second one to their *foundations*. More precisely, models (i) aim at reconstructing either network evolution processes or morphology; and, to that end, they (ii) rely on assumptions, or inputs, related again to either processes or morphology. This yields the double dichotomy shown on Table 1, which I will use as a guide for the remainder of this subsection.

2.1.1 Reconstructing processes

Let us first focus on generative processes at the lowest level i.e., the derivation of micro-level network generation rules governing the appearance or disappearance of nodes, and/or the formation or disruption of links (left side of Table 1).

Using micro-level processes. One of the most straightforward approaches consists in using, precisely, data describing these very dynamics, at the node and link level — abstracting low-level processes by observing low-level processes. In this category, we find what may be described as simple counting methods aimed at appraising the propensity of links to form preferentially more towards nodes possessing certain properties: the archetypal notion of "preferential attachment" (PA). In its most restrictive yet most widespread acceptation, PA relates to the ubiquitous observation that links tend to attach to nodes proportionally to their degree centrality (number of neighbors), or just degree. Following de Solla Price (1976), this acceptation essentially stems from Barabási et al. (2002). Several authors extended this notion beyond degrees to deal with a variety of both structural and non-structural features, including spatial distance (Yook et al., 2002), common acquaintances or topological distance (Kossinets and Watts, 2006), similarity (Menczer, 2004; Roth, 2005; Leskovec and Horvitz, 2008) or a combination thereof (Cointet and Roth, 2010). This approach may be followed the other way around: instead of descriptively measuring growth processes, normative processes may be proposed and compared against empirical link formation. For one, Papadopoulos et al. (2012) introduced a PA model based on a concept of geometric optimization: nodes are placed in a plane and new nodes may connect to a subset of existing nodes by minimizing a geometric quantity. The model thereby reproduces connection probabilities that match empirical observations in a selection of real networks, rather than infers such probabilities from real data.

Another set of approaches relies on machine learning: they principally aim at predicting the appearance of links by generalizing from past link creation. Scoring methods are among the simplest ones: for instance, Liben-Nowell and Kleinberg (2003) introduced a prediction function based on some dyadic feature (such as the

		Reconstructing: processes	Reconstructing: structure
Using	processes	Preferential attachment estimation, Link prediction, Classifiers, Scoring methods, ...	Preferential attachment-based generative models, Rewiring, Cost optimization, Social Simulation, Agent-Based Models (ABMs), ...
	structure	Exponential Random Graph Models (ERGMs), p_1, p^*, Markov Graphs, Stochastic Actor-Oriented Models (SAOMs), ...	Prescribed structure, Subgraph-based constraints, Kronecker graphs, Edge swaps, ...

Table 1: Double dichotomy of canonical network modeling approaches, which generally aim at reconstructing either evolution processes or network structure, and do so by relying either on evolution processes or network structure.

number of common neighbors, Jaccard coefficients, Katz distance). This function outputs a score on non-connected dyads for an empirical network observed over a learning period $[t_0, t]$. The prediction task then consists in going through the dyad list in descending order and comparing it with the links formed during a test period $[t, t']$. A large array of more sophisticated techniques have further been proposed, involving, inter alia, SVM classifiers (e.g., Adar et al., 2004) or more broadly supervised learning methods (Hasan et al., 2006), as well as matrix and tensor factorization (Acar et al., 2009) — see Lü and Zhou (2011) and Hasan and Zaki (2011) for introductory reviews of this type of endeavors.

On the whole, this strand of methods is rather geared towards prediction success rather than behavior estimation i.e., efficiently guessing which links will appear rather than providing explicit and interpretable link formation rules (see Yang et al., 2015, for a relatively recent discussion of their comparative performance). Besides, an increasing attention has lately been paid to the time-related and spatial variability of the prediction task by considering the local neighborhood of nodes, both in a topological and temporal manner (Sarkar et al., 2014) and in a semantic fashion e.g., by enriching the set of prediction features with content (Rowe et al., 2012) or so-called sentiment analysis (Yuan et al., 2014). The machine learning toolbox has finally been recently enriched with evolutionary algorithms (for instance, Bliss et al., 2014, evolve a weight matrix describing the relative contributions of various similarity measures in predicting new connections) and, more significantly, a wide array of network embedding techniques (Grover and Leskovec, 2016; Wang et al., 2016), which rely in part on the macro-level structure to create a multi-dimensional node representation space where the question of link prediction reduces to a geometric problem of efficient neighborhood exploration.

Using macro-level structure. Link formation principles may also be infered from the observed network topology. The most common approach here consists of statistical techniques which are not unlike what is typically done in econometrics. They aim at fitting a network-level model whose parameters are associated with specific link formation effects. In other words, such models rely on the network as a whole, rather than link sets.

Blockmodels could count as one of the simplest such techniques. Introduced in the 1970s in mathematical sociology, they typically apply to static network data. They rely on the assumption that the presence of links depends on latent "blocks" of actors who are loosely structurally equivalent (White et al., 1976) i.e., exhibit the same connectivity patterns toward other actors. Stochastic blockmodeling (SBM, Holland et al., 1983) proposes a probabilistic framework to statistically appraise inter-block connectivities (actors may also belong to a mixture of blocks, see Airoldi et al., 2008). In the same vein, some more recent works estimate the likelihood of missing links by using network divisions based on either dendrograms (Clauset et al., 2008) or blockmodels (Guimerà and Sales-Pardo, 2009).

Exponential Random Graph Models (ERGMs) famously belong to this class of approaches as well. In all generality, they rely on the assumption that the observed network has been randomly drawn from a distribution of graphs. The probability of appearance of a given graph is construed as a parameterization on a choice of typical network formation processes: be they structural (such as transitivity, reciprocity, balance, etc.) or non-structural (such as gender dissimilarity, homophily, etc.). The aim is generally to find parameters maximizing the likelihood of the observed network. Each parameter then describes the likely contribution of the corresponding category of link formation process (e.g., strong transitivity, weak reciprocity). ERGMs have been introduced by Holland and Leinhardt (1981) through the so-called p_1 model describing the probability of graph G as $p_1(G) \sim \exp(\sum_i \lambda_i v_i(G)) = \Pi_i \exp(\lambda_i v_i(G))$ where $v_i(G)$ denotes a value related to the i-th process (e.g., transitivity). p_1 assumes independence between dyads, which limits the model to simple dyad-centric observables: principally, degree and reciprocity. It can nonetheless be applied to a partition of the network into subgroups (Fienberg et al., 1985) or stochastic blockmodels (Anderson et al., 1992); parameters are thus a function of blocks. Frank and Strauss (1986) further introduced "Markov graphs", which take into account dependences between edges and thus triads and simple star structures, and which was subsequently extended as the p^* model (Wasserman and Pattison, 1996; Anderson et al., 1999; Robins et al., 2007). Generalizations to more complex graph structures have later been proposed e.g., for so-called "multi-level networks" (Wang et al., 2013; Brennecke and Rank, 2016), which are essentially graphs with two types of nodes and three types of links i.e., isomorphic to a socio-semantic network.

When longitudinal data is available, network evolution may also be construed as a stochastic process in itself. Holland and Leinhardt (1977) then Wasserman (1980) proposed to appraise network dynamics as a (continuous-time) Markov chain. They assumed that the probability of link appearance or disappearance depends on a limited set of (static) parameters representing the contribution of various structural effects, such as, again, reciprocity, degree; structure and behavior may be combined (Leenders, 1997). Networks observed at different points in time are used to fit these parameters. Albeit not directly affiliated with this framework, Powell et al. (2005) proceed in a similar fashion to determine the key factors guiding attachment of firms in a biotech sector. Stochastic actor-oriented models (SAOMs) further expand these ideas by introducing an actor-level viewpoint whereby actors establish link to optimize some objective function (Snijders, 2001). Again, the parameters of this function denote effects deemed important for link formation (or destruction). These models also accommodate some form of dyadic dependence, and take into account non-structural features such as gender. They may include behavioral observables (Snijders et al., 2007) or rely further on machine learning techniques e.g., by extending SAOMs to a Bayesian inference scheme (Koskinen and Snijders, 2007). In practice, SAOMs may be used to study non-structural effects linked to gender, racial, socioeconomic or geographical homophily, as demonstrated for instance in an online context on Facebook friendship (Lewis et al., 2012). ERGMs and SAOMs assuredly share several traits, and it is also possible to develop ERGMs in a longitudinal framework as temporal ERGMs (or TERGMs), where the estimation for a graph at time t depends on the graph at $t-1$ (Hanneke et al., 2010). For a more detailed comparison between SAOMs and ERGMs, see Block et al. (2019).

On the whole, these approaches enable the joint and concurrent appraisal of a variety of effects (each statistical model may consider an arbitrary number of variables to explain the shape of the observed network), with the drawback of reducing the description of the contribution of each effect to a scalar quantity.

2.1.2 Reconstructing structure

The second part of the double dichotomy (right side of Table 1) relates to the morphogenesis of the whole structure. It may again be roughly divided into two broad categories, depending on whether approaches are based on a given growth process or on the entire network topology.

Using micro-level processes. A myriad of models have been proposed to reconstruct network structure from normative assumptions. This is perhaps the most well-known and natural approach in statistical physics. At the core of these approaches lies generally a master equation or a master process featuring a certain number of key and oftentimes stylized ingredients. These ingredients correspond to an ideally small subset of canonical growth processes, defining the essential rules for adding –and, rarely removing– nodes, links, and most importantly specifying towards which types of nodes. The goal often consists in reproducing the observed connectivity (such as degree distributions), cohesiveness (such as clustering coefficients), or connectedness (such as component size distributions).

One of the earliest successful attempts at summarizing network morphogenesis with utterly simple processes consisted again in analytically solving simple PA based on node degree (Barabási and Albert, 1999). Models based on a general notion of PA have been extended in various directions: taking into account the age of nodes (Dorogovtsev and Mendes, 2000), their Euclidean distance (Yook et al., 2002), their intrinsic fitness (Caldarelli et al., 2002), their rank (Fortunato et al., 2006) or their activity (Perra et al., 2012), formalizing a notion of competition between nodes to attract new links (Fabrikant et al., 2002; Berger et al., 2004; D'Souza et al., 2007), copying links from "prototype" nodes (Kumar et al., 2000) or using random walks (Vázquez, 2003), introducing preferences for transitive closure (Holme and Kim, 2002) or for specific groups of nodes — based on an a priori taxonomy (Leskovec et al., 2005) or an affiliation network (Zheleva et al., 2009) — or mixing structural PA with semantic PA: for instance,Menczer (2004) introduces the so-called "degree-similarity" model after observing that connected web pages are rather more similar, while Roth (2006a) mixes group-based PA and semantic PA. Group-based PA may also be found in models based on teams, such as Guimera et al. (2005) whose network evolves through the iterative addition of network cliques between all team members, assuming a certain propensity to introduce newcomers and repeat past interactions. Socio-semantic team-based morphogenesis will be discussed in more detail in Section 3.3.

Another class of models is based on link rewiring. One of the simplest versions was introduced by Watts and Strogatz (1998), who start with a ring lattice of fixed degree and reconnect links with a given probability p. The resulting structure may be discussed in terms of low path length and high clustering coefficient, or "small-world". Colizza et al. (2004) later reproduced these two statistical features by adopting a distinct approach based on a rewiring process aimed at optimizing a global cost function, in a way inspired by Fabrikant et al. (2002).

Finally, a broad class of network models, especially in the social realm, falls into the category of *agent-based models* (ABM) as soon as they rely on a relatively rich combination of actor-centric processes. They generally aim at a specific application field which, in turn, requires detailed assumptions: as such, they typically offer a good combination of realism (they benefit from a stronger sociological grounding) and tractability (their study generally requires to resort to simulation). Examples of sophisticated models have been abundant in the social simulation literature from early on and are now present in a wide array of works at the interface between statistical physics and CSS. Let us nonetheless casually mention Pujol et al. (2005) who build various social exchange network shapes by combining various agent decision heuristics and cognitive constraints; and Goetz et al. (2009) who reproduce blogger posting behavior and citation networks by a combining random-walk-based generators and post selection rules. In section 4.3, I provide an overview of network-based ABMs targeted at the emergence of socio-semantic clusters, specifically polarization.

Using macro-level structure. Reconstructing graph structure directly from graph structure essentially means to show that some structural constraints entail other structural features — for instance by demonstrating that a certain number of connected components or a strong proportion of some sort of triads follows from a given degree or subgraph distribution. The earliest attempts precisely focused on the effect of prescribing a power-law degree sequence (Aiello et al., 2000) and, shortly thereafter, any degree sequence (Newman et al., 2001). Several methods have later been proposed for more sophisticated constraints, such as prescribed degree correlations (Mahadevan et al., 2006), subgraph distributions (Karrer and Newman, 2010), or recursive stuctures (Leskovec et al., 2010a).

A typical challenge consists in being able to sample the space of graphs induced by a given set of constraints. Some approaches manage to provide a closed-form expression of several average statistical properties of the induced graph space, as has been done for the typical path length or average clustering coefficient by Newman et al. (2002). When this is not possible, an alternative consists in sampling the graph space through iterative exploration: the initial empirical graph is typically transformed by swapping pairs of edges while respecting the original constraint (Rao et al., 1996; Gkantsidis et al., 2003). This corresponds to a navigation in a meta-graph gathering all graphs of the target space. Beyond simple constraints, exhaustive navigation is usually impossible. In Tabourier et al. (2011, 2017), I practically addressed this issue by introducing an empirical sampling method denoted as "k-edge switching", iteratively swapping groups of k links in order to cover an increasingly large portion of the underlying graph space.

2.1.3 Combining both: evolutionary models

In all four positions of the double dichotomy, the challenge generally consists in proposing some processes or constraints deemed to be key to explain network formation — be it transitivity, centrality, homophily, etc. The role of such and such mechanism may be either assumed *a priori*, then measuring how it contributes to generate such and such network shape, or measured *a posteriori*, by assuming its existence and appraising its magnitude during network evolution. In all cases, intuition is crucial: creating these models requires insights on which mechanisms may play a role. Yet, the role of some mechanisms and, more, their combination, may sometimes be counter-intuitive.

To alleviate this dependence on intuition, evolutionary algorithms were recently used to automatically infer candidate mechanisms from the observed structure. It differs from the above methods in that it jointly uses the structure to reconstruct processes and the processes to reconstruct the structure. More precisely, network structure is used to devise link formation processes and, in turn and iteratively, these discovered processes are precisely used to reconstruct the structure. Some of the earlier approaches introduced template models based on fixed sets of possible specific actions (e.g., creating a link, rewiring an edge, connecting to a random node, etc.). Actions may be organized in various manners: first as a fixed chart, resembling the typical structure of agent-based models (Menezes, 2011), as a sequential list of variable size (Bailey et al., 2012; Harrison et al., 2016) or, very recently, as a matrix whose weights describe the relative contribution of each action (Arora and Ventresca, 2017). In all these works, the evolutionary process aims at automatically filling the template model with actions and fitting the corresponding parameters. As is typical in evolutionary programming, it involves a fitness function to evaluate the resemblance between the empirical network and networks produced by the evolved model. Fitness functions rely on classical structural features (degree distributions, motifs, distance profiles, etc.). Models are iteratively evolved along increasing fitness values.

In parallel, we could contribute an original approach aimed at inferring arbitrarily complex combinations of elementary processes, construed as laws, embedded in a simple PA framework (Menezes and Roth, 2013, 2014). We first introduced a generic vocabulary making it possible to describe network evolution in a unified framework, as an iterative process based on the likelihood of appearance of a link between two nodes, construed as a function on node properties in the currently evolving network (i.e., a general form of PA). We used structural features such

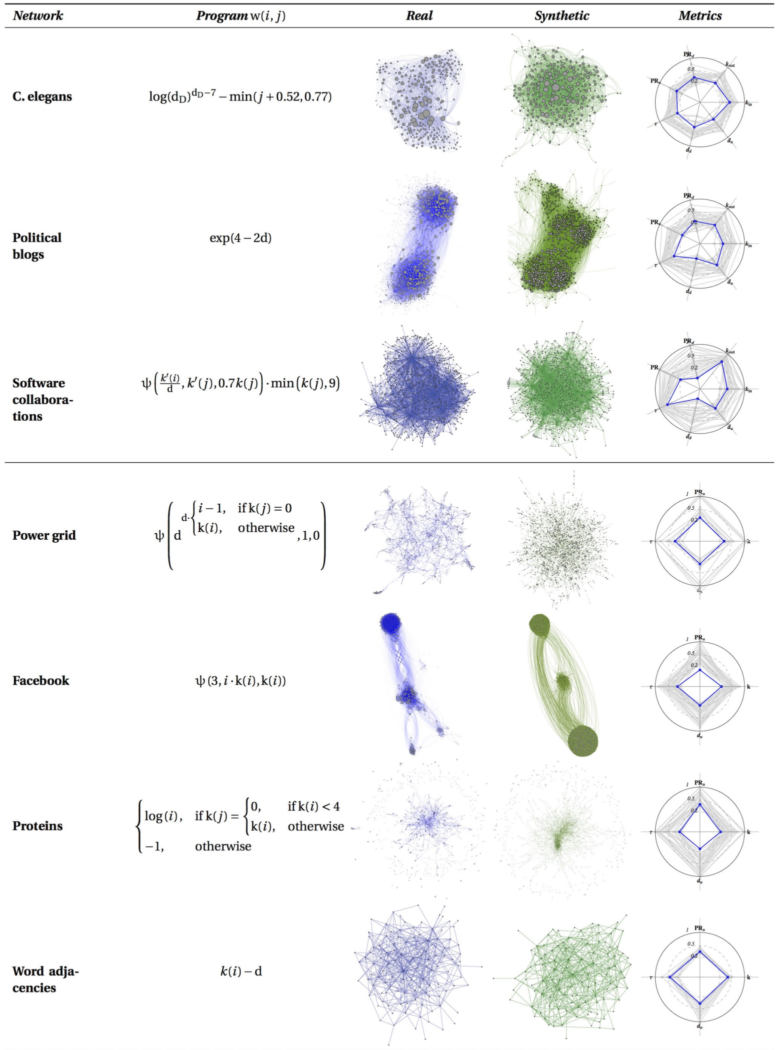


Figure 1: Empirical networks, automatically discovered rules, visual shapes of real networks and their synthetic equivalent, and radar describing the statistical accuracy of the reconstructed topologies. *From Menezes and Roth (2014).*

as distance, connectivity, as well as non-structural characteristics. Representing these functions as trees enabled us to apply genetic programming techniques to evolve rules which are then used to generate network morphologies increasingly similar to the target, empirical network. This technique may be denoted as "symbolic regression", for the goal is to genetically evolve free-form symbolic expressions rather than fitting parameters associated to fixed symbolic expressions: we *automatically* evolve realistic morphogenetic rules from a given instance of an empirical network, thereby symbolically regressing it. This strategy is inspired by Schmidt and Lipson (2009) who extract free-form scientific laws from experimental data. We first applied our method on kinship networks (Menezes and Roth, 2013) which led to a much more general framework in Menezes and Roth (2014). One remarkable result consists of the ability to systematically and exactly discover the laws of an Erdos-Rényi or Barabasi-Albert generative process from one of its stochastic realizations. Distinct, realistic and compact laws for a variety of social, physical and biological networks could also be found (see Fig. 1).

Families of network phenotypes and genotypes. Our approach provided a sort of artificial scientist proposing plausible network models, replacing the intuition of the modeler. It also makes it possible to discuss networks in terms of their plausibly genotype (i.e., generator equations), rather than phenotypes (i.e., a series of topological traits). Phenotypical traits assuredly provide the basis for appraising the quality of structural reconstruction and, by extension, for defining fitness functions attentive to such and such topological property (for an early yet already comprehensive review, see da Fontoura Costa et al., 2007). They also provide a good foundation for comparing networks with one another: a series of studies have indeed been devoted to defining network families by relying on triadic profiles (Milo et al., 2004), canonical analysis of various measures (da Fontoura Costa et al., 2007, §19), adjacency matrix spectrum (Estrada, 2007), blockmodeling (Guimerà et al., 2007), community structure (Onnela et al., 2012), hierarchical structure (Corominas-Murtra et al., 2013), communication efficiency (Goñi et al., 2013), graphlets (Yaveroglu et al., 2014), and graph embeddings (Yanardag and Vishwanathan, 2015). Phenotypical traits have also been the target of evolutionary algorithms in Märtens et al. (2017), who symbolically regress formulas describing the phenotype of the network e.g., expressing explicitly the diameter of various network classes as a function of the number of nodes, links, or some eigenvalues of the adjacency matrix.

By contrast, symbolic regression enables the comparison and categorization of networks based on their plausible underlying morphogenesis rules — as such a *genotypic categorization*. The research direction initiated in Menezes and Roth (2014) has further been developed in Menezes and Roth (2019a) to discover network families by categorizing their generators (genotypes) rather than their morphologies (phenotypes). These families are defined both in terms of the similarity of their function (i.e., how the generator works) and of their expression (i.e., its symbolic expression). To this end, we used 238 anonymized ego-centered networks of Facebook friends randomly sampled from about 10,000 such networks stemming from a large-scale online survey (under the collaborative ANR project "Algopol", specifically one of its experiments where consenting participants accepted to give access to their publication and network history; see also Bastard et al., 2017). For each network, we found a best generator expressed as a formula. To identify families of generators and to visualize how similar they are in relation to each other, we introduced a measure of dissimilarity between pairs of generators. To visualize the landscape of generators, we model these dissimilarities as distances in geometric space; their multi-dimensional scaling yields Fig. 2. This provides an overview of various generative processes for ego-centered networks, ranging from classical Erdős-Rényi graphs (the generator function is a constant c) or degree-based PA (generator function proportional to k), to generators which rely heavily on an affinity function (ψ) which corresponds to the existence of distinct classes of nodes; in other words, distinct social circles.

2.2 Unusual network topologies

I will present in more detail my contributions to graph morphogenesis in the context of socio-semantic networks in Section 3. Before that, I shall also evoke an area which is relatively peripheral in this regard, the morphogenesis of non-standard topologies of networks that are the object of some social science fields, yet are not social networks per se. Not all empirical networks are indeed sparse and heterogeneous. The opposite may be easily found in at least two specific fields, spatial networks (Barthélemy, 2011) and kinship networks (Hamberger et al., 2011), for which the understanding of morphogenetic dynamics warrants the development of distinct methods, and where I could also propose several contributions. Since this area is relatively further from socio-semantic issues, I briefly focus here on my most direct contributions.

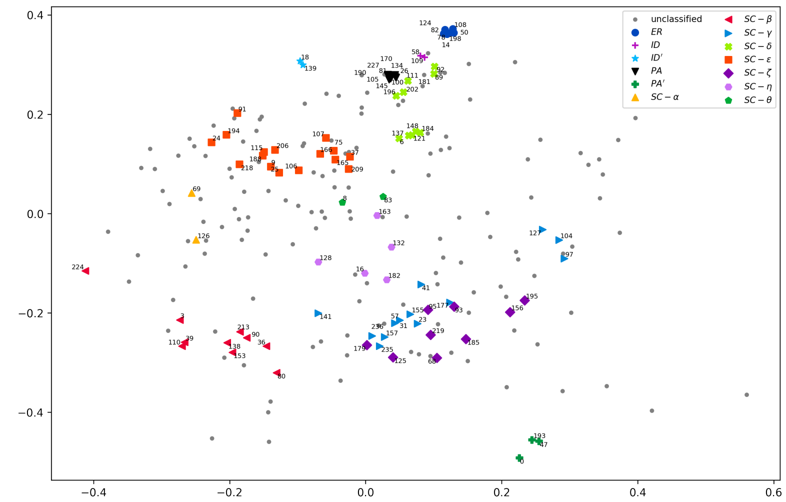


Figure 2: Network generators mapped on a two-dimensional layout according to their pairwise distances. Colors and shapes indicate generator families manually identified as semantically similar. ER means Erdos-Rényi (i.e., a uniformly random process reconstructs well the empirical network), PA means preferential attachment, ID means a very simple process proportional to node IDs. In particular, "SC" generators (for Social Circle) involve a notion of node class and are over-represented in Facebook friendship networks. *See Menezes and Roth (2019a)* for details.

Spatial and human mobility networks. The first strand of research focuses on the dynamics *on* transport networks i.e., human mobility. These networks often exhibit non-heteregeneous degree centrality distributions. They may either be close to complete networks: most nodes are connected to most other nodes, as is the case in spatial movement networks at a relatively short scale e.g., in an urban context where it is difficult to find two non-connected nodes). Or they possess a very modal degree distribution: most nodes have the same degree, as is the case in physical infrastructure networks. Of particular importance is the analysis of link weights with respect to space and physical distances.

Roth et al. (2011a) relied on a dozen of millions of subway rides made by a couple of millions of individuals within a large metropolis, London. Methodologically, the challenge consisted in describing patterns for an almost complete graph (all subway stations were connected with each other by at least one ride), while taking into account the geographical position of nodes. We could demonstrate that intra-urban movement is strongly heterogeneous in terms of volume, but not in terms of travel distance, thus debating previous results by González et al. (2008) and Brockmann et al. (2006). In both cases, we could show the existence of characteristic scales (contrarily to many real-world networks) and of a polycentric structure composed of large flows organized around a limited number of activity centers. For smaller flows, the pattern of connections becomes richer and more complex and is not strictly hierarchical since it mixes different levels consisting of different orders of magnitude. Furthermore, the spatial distribution of edges revealed a strong geographical heterogeneity and anisotropy; see Fig. 3.

Another strand of research relates to the dynamics *of* transport networks. In Roth et al. (2012a), I could show the existence of universal patterns of the evolution of the largest world subway networks along several decades, such as the existence of a very distinctive dichotomy between a dense two-dimensional core and a sparse, wide-reaching and one-dimensional periphery. Both network characteristics and geographical features could be described in simple ways in each regime (core/periphery). More recently, I could contribute to link transport network topology and to the underlying geographic-demographic characteristics of the underlying region (area, population, wealth), comparing urban and train networks (Louf et al., 2014).

A last strand of research relates to the description of mobility in open space, without considering transport networks. I could focus on the quantitative description of territory boundaries and, again, the existence of characteristic spatial scales. Human mobility is indeed known to be distributed across several orders of magnitude of physical distance, which makes it generally difficult to endogeneously find or define typical and meaningful scales. Relevant analyses seem to be relative to some ad-hoc scale, or free of any scale — be it for movement based on the circulation of artifacts (Brockmann et al., 2006), cell phone data (González et al., 2008) and calls (Sobolevsky et al., 2013), taxi rides (Liang et al., 2013), social media "check-ins" (Cho et al., 2011) or postings (Beiró et al., 2016). Network community detection algorithms used to find geographical partitions are similarly often based on either a single scale or ad-hoc scales (Thiemann et al., 2010; Sobolevsky et al., 2013).

In Menezes and Roth (2017), we demonstrated that mobility networks can enclose several coexisting and natural scales at the partition level, despite the scale-free nature of link distance distributions at the lower level. In other words, we automatically uncovered a small number of meaningful description scale ranges

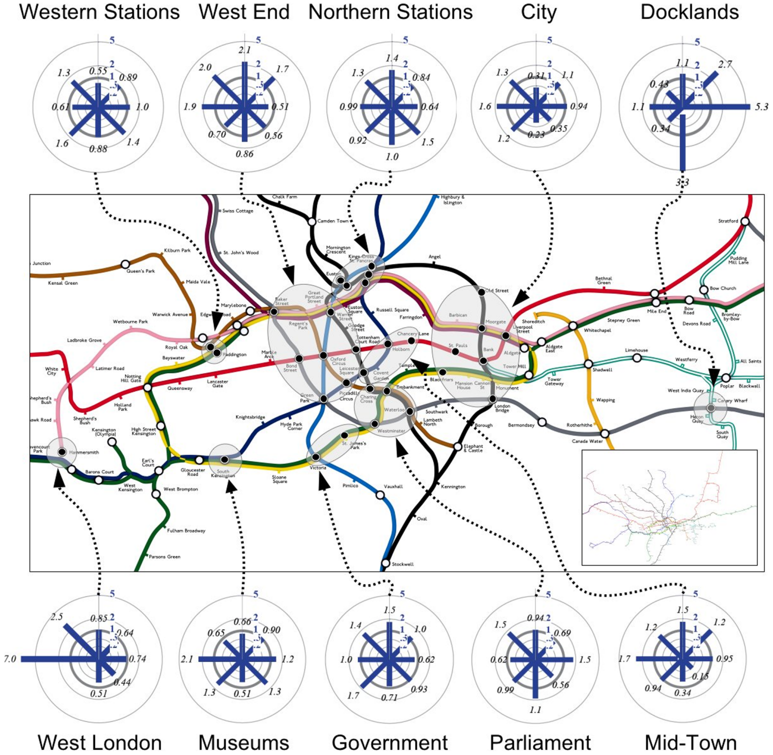


Figure 3: Polycenters and basins of attraction in the London subway system, representing the ten most important polycenters, and showing the corresponding propensity to anisotropy comparing actual flows with a null model only preserving total flows going to or from a given station (randomizing origin-destinations for that station); 1 means no deviation in a given direction. The anisotropy is essentially in opposite directions from the center, showing a strong bias towards the suburbs for peripheral centers essentially, rather than for central centers. Moreover, most stations control their own regions and seem to have their own distinctive basins of attraction. *From Roth et al. (2011a).*

from apparently scale-free raw data. We relied on geotagged data collected for a variety of geographical regions (from countries to metropolitan areas) from a photo-sharing platform, Instagram, over a period of 16 months. By tracking the places where a given user took photos we could infer the intensity of human movement between any two given locations in a region. We then defined a series of

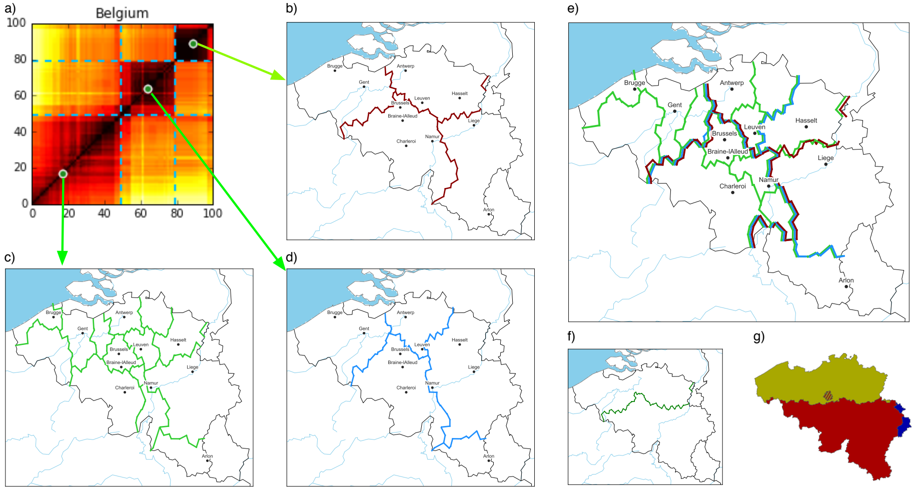


Figure 4: *Belgium borders at different scales.* a) Heat map of scale similarities 1; b) Borders for the long distance scale; c) Borders for the short distance scale; d) Borders for the middle distance scale; e) Multiscale borders; f) Borders based on optimal two community partition of the full graph; g) Language communities of Belgium. *From Menezes and Roth (2017).*

movement networks constrained by increasing percentiles of the distance distribution, to which we apply a relatively straightforward community detection process. Using a simple parameter-free discontinuity detection algorithm, we discover clear phase transitions in the community partition space. Fig. 4a focuses on the case of Belgium, further illustrating how natural scales correspond to partitions in the map, and how the several natural scales can be combined in a single multiscale map, which provides richer information about the geographical patterns of the region than is possible with more traditional methods. Further, the analysis of scale-dependent user behavior hints at scale-related behaviors rather than scale-related users: a core of users are active in all scales. Besides, we showed that the ambition of finding natural phases in community partitions based on some notion of resolution, which can be fulfilled in non-geographical scale-free networks (Traag et al., 2013), could also be tackled in the case of spatial mobility networks. More broadly, this allows the introduction of boundary conditions based on a scaffolding of a small number of natural spatial and behavioral scales emerging endogenously from the data, which provide a subset of relevant description levels.

Kinship networks. Kinship anthropology deals with graphs variously describing genealogical relationships or alliance networks. The former are strongly constrained graphs, featuring two ascendants per node and temporal order. The latter are weighted and, often, non-sparse directed graphs describing marriages between well-defined kinship groups (see Fig. 5). In both cases, the challenge is to exhibit regularities or anomalies in matrimonial strategies to quantitatively falsify established anthropological theories, especially as regards relinking phenomena (e.g. two brothers marrying two sisters of another family, thus creating a cycle in the genealogical network, or a circuit in the alliance networks). While coordinating the CNRS partner of the "SIMPA" ANR project (for an overview, see Menezes et al., 2016), I could in particular tackle a variety of combinatorially challenging issues for such networks, including the counting of cycles and development of multinomial models for weighted dense networks (Roth et al., 2013), the automatic discovery of matrimonial rules, successfully applying the above-mentioned evolutionary models to kinship and genealogical newtorks (Menezes and Roth, 2013; Menezes et al., 2016).

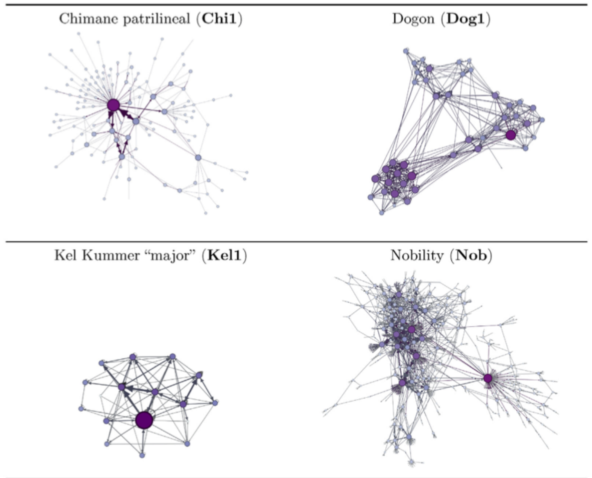


Figure 5: Some empirical alliance networks: node size is proportional to strength, while arrow width is proportional to arc weight (graphs were rendered with GePhi using the ForceAtlas visualization algorithm). *From Roth et al. (2013).*

3 Socio-semantic formalisms

Social cognition may be studied at the junction of these disciplinary efforts. Networks offer a practical and fertile perspective to appraise jointly the interactional substrate and the cognitive landscape within a similar formalism. Albeit obeying distinct processes, the morphogenesis of social and semantic networks may be formulated in a unified framework.

In my work, the idea of binding the social and semantic realms first appeared in a preprint (Roth and Bourgine, 2003) before introducing explicitly the label "socio-semantic network" (Roth, 2005) and formalizing this further as "epistemic networks" (Roth, 2006a), leading to a family of structures related to graphs (social and semantic), bipartite graphs (socio-semantic), hypergraphs (with hybrid social and semantic hyperedges) and lattices (hybrid taxonomies of actors and concepts) (Roth, 2013). As stated before, a string of works emphasized the importance of intertwining content and social structure — in SNA (Emirbayer and Goodwin, 1994), lattice analysis (Mische and Pattison, 2000), visualization (Sack, 2003) or the semantic web (Mika, 2005) — yet it is only towards the end of the 2000s that this combination appeared in a significantly increasing number of empirical studies (Adamic and Glance, 2005; Etling et al., 2010; Wang and Groth, 2010; Conover et al., 2011; Lietz et al., 2014; Martin-Borregon et al., 2014; Ciampaglia et al., 2014; Korff et al., 2015; Himelboim et al., 2017, to cite a very few).

This binding is also behind several projects which I could supervise at the interface between computer science and sociology, especially a series of grants successively addressing the morphogenesis, diffusion patterns and informational diversity of online communities and the digital public space (ANR Webfluence, 2009-10; ANR Algopol, 2012-15; ANR Algodiv, 2016-19; ERC Socsemics 2018-23). In parallel, I developed developed some epistemological reflections on the modeling of socio-technical or socio-semantic systems in a series of subsequent articles (Roth, 2007a, 2009, 2010, 2013), while pursuing the empirical applications of such frameworks (including Tamine et al., 2016; Roth, 2017; Baltzer et al., 2019; Roth and Basov, 2020; Medeuov et al., 2021). This section aims at discussing how socio-semantic frameworks have been progressively introduced to understand social cognition processes at the micro-, meso- and macro-levels. Formally, this respectively corresponds to the notions of socio-semantic graphs, hypergraphs and lattices.

3.1 Socio-semantic ontology

Ontological difficulty and assymmetry. One of the first issues with socio-semantic frameworks consists of a fundamental asymmetry in the burden of defining what counts as an entity in each realm. On the social side, it appears considerably easier, if not straightforward in some contexts, than on the semantic side. Social nodes typically represent individuals (such as authors, users), rarely collective entities (such as organizations, websites) whose boundaries are admittedly less well-defined. Semantic nodes and, thus, semantic categories, pose significantly more definitional questions, especially as they are often abstracted from text corpuses (Evans and Aceves, 2016), be it interview transcripts or digital publications. Shall we focus on words, n-grams, or terms? This goes further than the mere technical processing questions evoked earlier in Section 1.3, as word meaning across different contexts and, more broadly, different pieces of text is not obviously stable (Leydesdorff, 1997). Even if we dismiss this problem, shall we use terms, or "topics" as term sets or distributions (Grimmer and Stewart, 2013)? "Statements" or "clauses" as syntactically articulated terms (Mausam et al., 2012; Vossen et al., 2016; Van Atteveldt et al., 2017)? Whole "sentences" as syntactically consistent units (Leskovec et al., 2009)? Vectors in some opinion or epistemic space (Gilbert, 1997; Sim et al., 2013)? Vector models (Salton et al., 1975) up to the most recent machine learning (ML) approaches aimed at geometrically embeddinging words, sentences or documents, or both (respectively Mikolov et al., 2013; Le and Mikolov, 2014; Angelov, 2020, which, to some extent, take implicitly into account an underlying relational and contextual structure) ? Answers and decisions considerably vary across the state of the art.

The definition of links easily generates further debates. On the social side, they are admittedly an abstraction of reality (Fuhse, 2009), all the more when the same data induces various types of relationships (affiliation vs. interaction; retweets vs. mentions, see Conover et al., 2011; Lietz et al., 2014) or layers (Kivelä et al., 2014). Deciding how to attribute weights to edges, or arcs, and why, is not easier. Such discussions nevertheless remain generally short: binarization is very frequent and weights oftentimes correspond to interaction frequency, whatever the interaction. On the semantic side, link meaning is relatively obvious when nodes are terms (co-occurrence frequency is a consensual choice, while SNA techniques may also indifferently be applied on semantic networks, as in e.g., Carley, 1986; Nerghes et al., 2015) — and much less otherwise: a myriad of similarity metrics are available (Jaccard, cosine, edit distances in some space) to express the relatedness of words, sentences or documents. Last but not least, there is no less freedom for defining connections between social and semantic entities and most of these issues remain equally acute.

The formalism of **socio-semantic hypergraphs** remains generally agnostic of these issues. It accommodates most of them by allowing a variety of assumptions on the definition of nodes and relationships.

Most importantly, socio-semantic hypergraphs accommodate most of the operational frameworks of the literature reviewed in the present manuscript.

This formalism distinguishes two sets of nodes respectively representing social (S) and semantic entities ($\mathcal{S}$). It configures a single type of n-adic relation X, which is a (possibly valued) subset of the hybrid powerset $\mathcal{P}(S \cup \mathcal{S})$. See Fig. 6 for a generic illustration.

Most digraphs used to empirically connect social and semantic structures may be construed as some restriction of X over some set product defined over S and $\mathcal{S}$. Consider for instance the structure of Fig. 7 as a socio-semantic hypergraph. The dyadic social relation $s = X \cap (S \times S)$ defines a classical binary social network. Similarly, $\mathscr{s} = X \cap (\mathcal{S} \times \mathcal{S})$ induces a classical binary semantic network. Socio-semantic relationships between social and semantic entities may be defined as $\chi = X \cap (S \times \mathcal{S})$, and so on — the generalization to directed, weighted or timestamped relations X is straightforward (using directed, set-valued hypergraphs), albeit normally beyond the scope of this section.

The so-called "socio-semantic networks" of Roth and Bourgine (2003), Roth (2013), and Roth and Basov (2020), as well as what I initially called "epistemic networks" in Roth (2006a, 2007b,c, 2008a) are precisely given by these s, $\mathscr{s}$ and χ. By contrast, Roth (2005), Cointet and Roth (2009, 2010), Roth and Cointet (2010), and Baltzer et al. (2019) only make use of s and χ (which are also both implicitly present in Menezes et al., 2010a). Finally, Roth and Bourgine (2005, 2006), Roth et al. (2008a), and Roth (2008b, 2010, 2017) just use the more classical bipartite relationship χ, which I may also sometimes have (confusingly) denoted as a socio-semantic network even if it is actually restricted to dyadic links between exactly a social and a semantic entity: to avoid any ambiguity, I shall from now on denote this bipartite relationship as the "socsem network". Likewise, links between social and semantic entities are denoted as "socsem links".

Socio-semantic hypergraphs cover many other usual uses. For instance, X may describe hybrid n-adic affiliations (collaborations, co-attendance, co-membership gathering any number of actors and concepts) and were introduced as such in

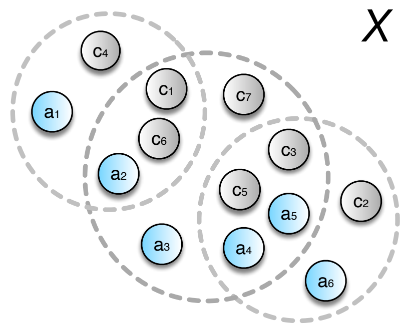


Figure 6: Toy example of a socio-semantic hypergraph X. Nodes represent either agents from S or concepts from $\mathcal{S}$. The boundaries of three partially-overlapping socio-semantic hyperedges are figured by thick dashes: the top-left hyperedge, for instance, gathers $\{a_1, a_2, c_1, c_4, c_6\}$.

Taramasco et al. (2010). In this case, the classical projected social network may also be written as $S = \{(a, a') \in S^2 \,|\, \exists x \in X, (a, a') \subseteq x\}$. The classical co-affiliation network may be defined dually as $\mathcal{S} = \{(c, c') \in \mathcal{S}^2 \,|\, \exists x \in X, (c, c') \subseteq x\}$, just as the affiliation relation χ is defined analogously as the projection of X on $S \times \mathcal{S}$. Considering s, $\mathscr{s}$ and χ as adjacency matrices isomorphic to the relations they each represent, we can write $s = \chi\chi^T$ and $\mathscr{s} = \chi^T\chi$, which reveals the usual correlation between these three dyadic networks (Everett and Borgatti, 2013).

Semantic networks centered on actors, for instance on a given actor $a \in S$, are induced by $\mathscr{s}_a = X \cap (\{a\} \times \mathcal{S} \times \mathcal{S})$ if X is defined such that $(a, c, c') \in X$ indicates that the connection between concepts $(c, c') \in \mathcal{S}^2$ holds for a i.e., is in the semantic network of a (used in Medeuov et al., 2021). This covers the issue of distinct actors ascribing different semantic relationships between terms, and even different meanings (Fuhse, 2009). It enables the definition of multi-actor-multi-concept patterns, and the comparison between actors of their respective semantic networks. For instance, socio-semantic diamonds of Roth (2006a) and Roth and Cointet (2010) describe the joint use by a and a' of c and c' and correspond to the case where (a, a', c, c') belongs to some hyperedge of X. These hybrid patterns are currently at the root of a burgeoning literature building in part upon the above socio-semantic formalisms (see e.g., Basov and Brennecke, 2017; Basov et al., 2019, 2020; Fuhse et al., 2020).

"Socio-semantic" et al. Cultural sociology in particular (Section 1.2) has developed formal frameworks that connect, to some extent, social and semantic networks, or semantics with social networks, or semantic networks with actors (most notably, to recall a few, Carley, 1986; Leenders, 1997; Mohr and Duquenne, 1997; Lazega and Pattison, 1999; Snijders et al., 2007; Vaisey and Lizardo, 2010; Pachucki and Breiger, 2010; Godart and White, 2010; Wang et al., 2013). Few works pushed the generalization of connection types much further than Krackhardt and Carley (1998) who introduced networks between several kinds of entities (actors, resources, tasks, skills) and their various potential projections, denoted as "meta network" in Carley (2002) and leading to a further operationalization in Diesner and Carley (2005) where a myriad of, inter alia, actor-actor, actor-concept, concept-concept relationships are extracted from text corpuses.

Notwithstanding, a few fields have explicitly coined new notions and expressions that combine "social" and "semantic" term. Such is obviously the case of the "*socio-semantic web*" or "*social semantic web*" (Zacklad et al., 2003; Wang and Groth, 2010), aimed at integrating social aspects into the so-called "semantic web" which, at the time, was an already-thriving engineering field focused on ontologies, and precisely on the conception of specifications and protocols to deal with relationships between various types of entities. The socio-semantic web intended to augment semantic web ontologies with a social layer, and the social web (i.e., essentially ontologies of people) with a semantic layer, in order to improve the cooperative management of ontologies and to foster the emergence of knowledge-oriented user groups. While the so-called "FOAF" (Friend Of A Friend) ontology connects almost indifferently actors and concepts in all directions (Finin et al., 2005), by contrast, "*semantic-social networks*" (Mika, 2007) extended classical ontologies by adding the notion that meaning is agent-dependent i.e., specifying which actor is behind which ontological relationship. Such tripartite actor-concept-instance graphs resemble very much both social tagging systems and the precursor "*cognitive social structures*" proposed by Krackhardt (1987): the resulting macro-level structure is essentially a knowledge map whose implicit connectors are agents or, more precisely, their beliefs on concepts. On the whole, these efforts nonetheless tend to bring to web services a social epistemology perspective (1.1) rather than a cultural sociology one (1.2).

This notion of collaborative knowledge construction also applies particularly well to science models (Payette, 2012). The seminal paper by Gilbert (1997) features "kenes", a two-dimensional knowledge vector analogous to genes, and a simple dynamics where, is essence, the generation of new papers by (implicit) scientific actors is influenced by the kene-based position of previous papers, which remarkably reproduces realistic connectivity and cluster structures. This further paved the way for "*epistemic networks*" whose actors pursue a goal of scientific or epistemic exploration in some well-defined knowledge space. Typical questions relate to the effect of the social structure on the efficiency of the semantic exploration (Grim, 2009; Weisberg and Muldoon, 2009) and, dually, to the influence of the shape of the semantic space on interaction processes (Sun et al., 2013).

Hybrid networks of actors and concepts, featuring in all generality hybrid interactional and semantic relationships, were increasingly introduced around the turn of the 2010s. They were variously denoted as "*semantic social networks*" (Gloor et al., 2009), "*attribute-augmented graphs*" (Zhou et al., 2009), "*bi-type information networks*" (Sun et al., 2009), "*content-based social networks*" (Cucchiarelli et al., 2010), "*social content networks*" (Wang and Groth, 2010), "*augmented social networks*" (Cruz et al., 2013), or "*heterogeneous graphs*" (Liu et al., 2017), to cite a few. These works typically involve both social and semantic nodes, and describe links and clusters of interest in a unified representation. This also relates to a longer effort in the information visualization community to propose hybrid social and semantic representations, starting with Sack (2003) who put side by side semantic and networks (term relationships and user replies), responding to the short yet much older call of Dourish and Chalmers (1994) to do so. Hybrid visualizations have since become much more familiar, including "pivot graphs" (Wattenberg, 2006), where node placement is influenced by attributes, "pivot paths" (Dörk et al., 2012), which alternates between actor- and concept-centric navigation. Various types of projections of social graphs on semantic maps (Korff et al., 2015) or the other way around (Gaumont et al., 2018) are now rather common.

3.2 Micro-level socio-semantics: nodes & links

All these frameworks support the appraisal of socio-semantic morphogenesis. Let us focus for now the micro level i.e., the correlations between intertwined structural and semantic measures at the level of actors and concepts, both in a static and diachronic way. The formalism of choice, here, is the above-mentioned socio-semantic network, *as a graph* featuring strictly dyadic connections and defined by s, $\mathfrak{s}$ and χ (Fig. 7).

This first concerns static characterizations and, more precisely, a posteriori observations of socio-semantic correlations. Most studies in this field compare structural measures on s and on χ, thereby showing that social and socsem degree centralities (akin to social and cultural capitals) are both heterogeneous (many have little, few have a lot), strongly correlated, and that these features are stable across time in spite of a strong low-level activity and thus vigorous network evolution (as in Roth, 2006a; Roth and Cointet, 2010, see also Fig. 8a-top) while they jointly plateau (Baltzer et al., 2019), suggesting correlated underlying constraints. Similarly, there is socio-semantic assortativity in the sense that connected nodes in the social network are also socsem-connected to similar semantic entities (Aiello et al., 2010; Conover et al., 2011; Mitzlaff et al., 2013; Lietz et al., 2014; Cinelli et al., 2021). More sophisticated socio-semantic correlation patterns could also be exhibited: for one, the classical notion of "structural holes" (Burt, 2004), denoting the brokerage position of certain nodes in a social network which are connected to otherwise weakly connected portions, could be translated in a socsem context as "cultural holes" (Vilhena et al., 2014). Such holes describe divergences in the way actors use terms: the underlying idea is that communications should be more difficult when socsem connections match less i.e., are holes in the socsem network. This leads to double dichotomies connecting various levels of structural embeddedness with various levels of "cultural" embeddedness, while demonstrating a non-monotonous relationship between the two. Simply put, cultural brokers are not necessarily strongly connected socially, and vice-versa (Goldberg et al., 2016; Garimella et al., 2018).

Socio-semantic evolution. Second, several studies examined low-level socio-semantic processes in a diachronic or longitudinal manner, in order to show how socio-semantic properties in the broad sense may a priori influence link formation. Some analytical network formation models (Boguna and Pastor-Satorras, 2003; Boguna et al., 2004) assume a joint influence of latent socsem connections in χ on the appearance of social links in s. Empirically, this is at the root of Roth (2006a), Cointet and Roth (2009), and Roth and Cointet (2010) which address this question in a co-evolutionary framework in both scientific communities (authors, concepts) and blog networks (bloggers, topics). An important finding is that the relationship between social link formation likelihood and socsem distance (denoted as δ) measured either as a Jaccard coefficient or as a cosine distance on the concept sets that actor dyads are respectively using, appears to be non-

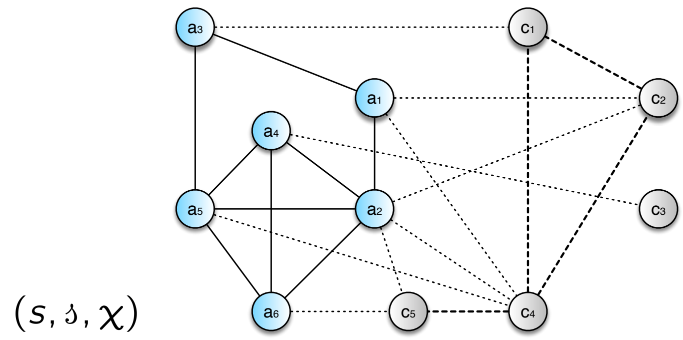
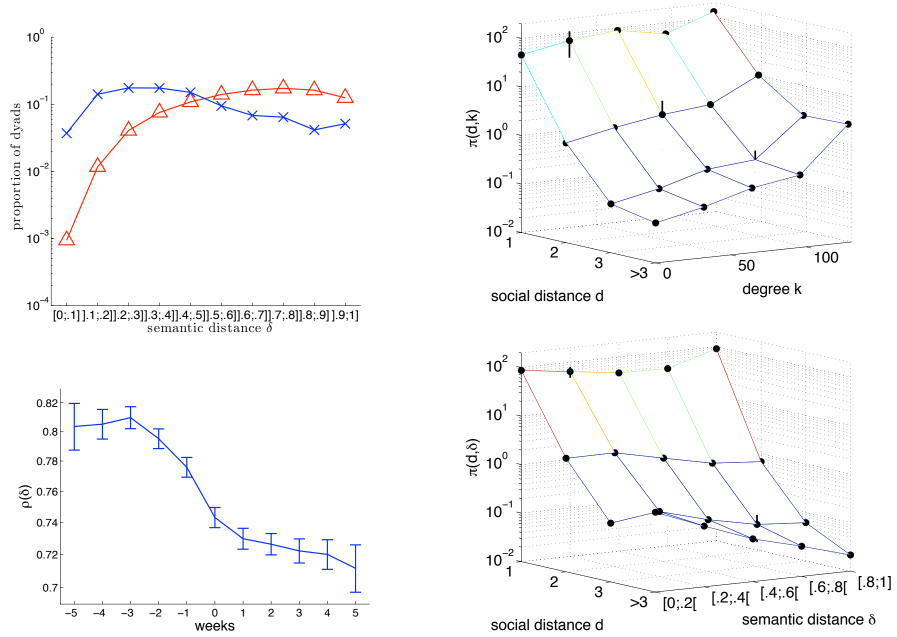


Figure 7: Ontology of a typical socio-semantic network given by the triplet $(s, \mathfrak{s}, \chi)$ i.e., social, semantic, and socsem networks: nodes $a \in S$ denote actors, while nodes $c \in \mathcal{S}$ denote concepts; links correspond diversely to interaction, co-occurrence, or usage; they may be directed or undirected. This socio-semantic network is not derived from the example of Fig. 6 and illustrates the situation where links of s, $\mathfrak{s}$ or χ do not configure cliques.

monotonous in interaction networks (collaboration) while it is monotonous in affiliation networks (citation), suggesting that the former is less homophilic than the latter. Several of these findings could be replicated on Twitter by Šćepanović et al. (2017) while additionally focusing on link deletion. Further, the socio-semantic framework sheds a different light on classical reinforcement processes (of the "rich-get-richer" type). As is traditionally the case, more connected agents benefit proportionally from new connections, and agents who are topologically closer in the social network are exponentially more likely to attract new connections (topological distance d considerably matters). In passing and interestingly, social capital appears to matter more when social distance is higher: in the local neighborhood of repeated or "friend-of-friend" interactions, capital (or, perhaps, fame) matters less. The influence of social topology is further intertwined with socsem similarity in a non-trivial manner: we observed in Cointet and Roth (2010) that interaction propensity, which is smaller toward socially remote actors, also increases with socsem similarity whenever agents have not interacted before (i.e. for $d > 1$); whereas this is much less so for repeated interactions ($d = 1$) where socsem homophily plays a weaker role. The bottom graph of Fig. 8b describes the magnitude of the interaction propensity with respect to both social distance d and semantic distance δ. On the whole, reinforcing processes may be characterized in all directions for all social, socsem and semantic networks: Wang and Groth (2010) appraised the strength of the aggregated influence of each network on each other, for several classical network properties, in a unified framework that relies exactly on a socio-semantic network formalism; for instance, they show that clustering in the social network has a ".55" positive influence on centrality degree in the semantic network.

The above phenomena also point to the existence of two types of interaction modes, local and distant, featuring distinct socio-semantic processes, which makes it possible to speak of a "local" rather than "small" world, defined as a small circle around ego where link formation happens in a distinct manner than for long-distance links. This further hints at two other phenomena: socio-semantic contraction and dependence on latent groups. Regarding the former, not only social or socsem proximity induces a higher likelihood of connection (socsem selection), not only does it increase after interaction (socsem influence), but it does also appear to increase before interaction: both social and socsem proximity obey a sigmoid function whose characteristic shape starts several periods of time before actual first interaction, in contexts as varied as user-to-user interaction on Wikipedia discussion pages (Crandall et al., 2008) or blog networks (Cointet and Roth, 2010, see Fig. 8a-bottom). Regarding the latter, the presence of underlying groups that may also further influence the morphogenesis: Zhou et al. (2006b) study both the socsem adoption of concepts (how likely an actor a using c is to use c') and the joint social & socsem adoption (how likely an actor a' connected to a who uses c is to use c') and demonstrate the existence of a block structure in concept adoption: put differently, some subsets of concepts are likely to induce the socsem- or socially-mediated adoption of some other subsets of concepts. This hints at the meso-level effects that will be the focus of the next subsection.

Concept adoption may here be regarded as the formation of socsem links that depend on the socsem and social structure as a whole. This process has been exhibited in the case of the longitudinal adoption of values or world views dependent on prior connections between actors (Vaisey and Lizardo, 2010), whereby the socsem neighborhood of connected actors is becoming more similar across time; or in the longitudinal convergence of discourse in an organizational setting (Saint-Charles and Mongeau, 2018) whereby the same coevolutionary phenomenon is observed, and socsem similarity is also correlated with social centrality, albeit in a non-monotonous way (above a certain threshold, socsem similarity does not



(a) (b)

Figure 8: **(a)** Semantic proximity of connected dyads: *top*, comparison between the semantic dissimilarity of connected dyads (blue crosses) and pairs of nodes in the whole network (red triangles); *bottom*, evolution of the average relative semantic dissimilarity $\rho(\delta)$ over all pairs of nodes getting connected for the first time at week 0. **(b)** Joint socio-semantic interaction propensity with respect to topological distance d and social capital k (*top*) or semantic dissimilarity δ (*bottom*). *From Cointet and Roth (2010).*

induce reinforcement dynamics in terms of social centrality); or in the adoption likelihood of cultural elements in the context of fashion (Godart and Galunic, 2019), which is higher when concepts are more strongly embedded in the semantic network (relying here on $\mathcal{S}$) and less popular in the socsem network (χ), translating the fact that fashionable elements should both be well inserted in the semantic structure and not too mainstream in the socsem structure.

Socio-semantic processes are also the relatively recent target of ERGM-based studies which consider so-called "multi-level" (Wang et al., 2013) and "multiplex" patterns (Basov and Brennecke, 2017). These patterns are expressed in a framework isomorphic to socio-semantic networks and serve as the basis for the appraisal of, say, the contribution of (a, a', c, c') motifs to socio-semantic morphogenesis. The very recent work by Koskinen et al. (2020) interestingly considers groupings of semantic links as nodes of a further semantic hypernetwork which, for one, does not conform to the socio-semantic hypergraph formalism because of its recursivity.[1] Let us finally mention some ML techniques of link prediction in socio-semantic networks, for instance by including socsem similarity into the scoring of social links (Hours et al., 2016) or by using network embedding such as semantic proximity search in "heteregeneous graphs" (Liu et al., 2017). Efficiency is, again, the main target of ML methods, generally at the price of a lower explanatory power.

3.3 Meso-level socio-semantics: collectives & hyperedges

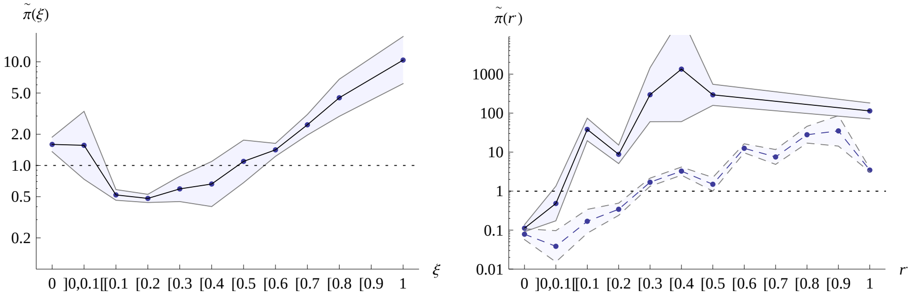
3.3.1 Collectives at the micro-meso level: socio-semantic teams

Several socio-semantic systems, such as science or open-source software development, feature horizontal, self-organized team work. Individuals more or less freely decide to gather in teams to produce knowledge: social cognition occurs not only at the macro-level of the whole system, but also at the meso-level of teams, which ideally aspire to achieve the best possible mix of skills by optimizing social and cognitive affinities of the group. We touch here the limits of dyadic network frameworks: team processes are not a sum of individual rationalities and some characteristics may not be expressed at the dyadic level. The seminal study of Ruef (2002) showed how several factors including gender, status, or ethnicity, influence the propensity to compose a team of company founders. Several subsequent studies described the configuration and performance of knowledge creation teams by focusing essentially on the social structure (Uzzi and Spiro, 2005; Jones et al., 2008) or on a few attributes (typically gender or ethnicity, see Zhu et al., 2013; AlShebli et al., 2018) — denoting a broader and increasing interest in what has meanwhile become "team science" (Stokols et al., 2008; Ramos-Villagrasa et al., 2018).

[1] This hypergraphic recursivity is also present in what we introduced as "semantic hypergraphs" in Menezes and Roth (2019b) to deal with sophisticated semantic and linguistic constructs per se. As such, they have more to do with computational linguistics and are thus beyond the scope of the present manuscript.

Hypergraphs are a natural modeling framework for teams. They appropriately generalize graphs: hyperedges gather an arbitrary number of nodes and not just two, by design (graphs are indeed hypergraphs whose hyperedges are of fixed size two). They have been used relatively early in categorization tasks (Gibson et al., 2000): for instance, Zhou et al. (2006a) showed their superior efficiency in classifying terms, while Sharma et al. (2014) developed a prediction framework for hypergraphic social collaborations. Hypergraphs nonetheless appeared sporadically over the last two decades, until they started generating very recently a fast-growing literature in the network science community (Battiston et al., 2020). However, most of this research strand principally relies, at least implicitly, on a perfect isomorphism between bipartite graphs and social hypergraphs: one side consists of actors, the other denotes teams or groups (as in, actually, Breiger, 1974) i.e., equivalently, social hyperedges. By contrast, socio-semantic hypergraphs are normally not isomorphic to some social hypergraph or bipartite graph.

Teamwork depends on cognitive properties: teams are formed according to both social and semantic features; socio-semantic hypergraphs further appear as a relevant description level. While the team-based nature of academic collaboration has long been underlined (deB. Beaver, 1986), quantitative and formal frameworks have traditionally been based on graphs (Mullins, 1972; Newman, 2001) or multi-dimensional surveys (Cummings and Kiesler, 2007). Hypergraphs appear sometimes explicitly in the scientometric literature, yet to characterize actor-centric properties (Han et al., 2009; Lung et al., 2018). In Taramasco et al. (2010), to my knowledge for the first time, we appraised academic teams as a socio-semantic hypergraph X based on $S \cup \mathcal{S}$ (see Fig. 6). Let us present its key points. The empirical data stemmed from bibliographical records which originally conform the hypergraphic ontology: papers gather authors (actors) and concepts (e.g., lemmatized salient terms extracted from abstracts); in practice, we selected some fields (focused on topics: rabies, the zebrafish model animal, or on actors: FAO/WHO expert groups) gathering thousands of papers and tens of thousands of articles over a couple of decades (1985–2007). This defines an empirical dynamic hypergraph X_t that grows through the cumulative addition of hyperedges x, each describing authors and scientific concepts which participated in the same paper. X_t gathers all papers until year t (drawn as "history graphs" by Datta et al., 2014). Simple hypergraphic measures may be defined, at any time, depending on past team arrangement. For instance, the *socio-semantic expertise ratio* of a hyperedge x in a given concept c at time t, noted $\xi_{c,t}(x)$, denotes



(a) **(b)**

Figure 9: Hypergraphic propensity (team bias): **(a)** for the proportion of experts per article; **(b)** for hypergraphic repetition ratios (social in solid line, semantic in dashed line). (Geometrically averaged behavior based on all four datasets.)

the number of actors of x who already appeared in at least one past hyperedge containing c before t (we call such actors "experts" in c). That is,

$$\xi_{c,t}(x) = \left|\{a \in x \cap S \mid \exists x' \in X_{t-1},\ \{a, c\} \subseteq x'\}\right| / \left|\{x \cap S\}\right|$$

Going further, the degree of social originality of a hyperedge x at time t, or *social hypergraphic repetition ratio* $r_t^S(x)$, may be measured by counting the proportion of subsets of actors of x which were already included in a past hyperedge at $t' < t$: it goes from 1 when all actors were previously all together in at least one collaboration, to 0 when the team does not even feature a single pair of previously interacting scientists. In formal terms,

$$r_t^S(x) = \left|\{x' \in \mathscr{P}(x \cap S) \mid \exists x'' \in X_{t-1}, x' \subseteq x''\}\right| / \left|\mathscr{P}(x \cap S)\right|$$

Conceptual originality may be dually measured by a *semantic hypergraphic repetition ratio* $r_t^{\mathcal{S}}$ based on concepts i.e., replacing S with $\mathcal{S}$ in the previous formula.

These indices make it possible to describe the *a posteriori* composition of teams, in terms of the raw distribution of teams exhibiting a given expertise ratio ξ or given social or semantic hypergraphic repetition ratios r^S or $r^{\mathcal{S}}$. We first observed that social originality is lower for extreme values of ξ i.e., for teams made of experts only or made of non-experts only: hyperedges of a mixed level of expertise correspond on average to a more original gathering of individuals. However, there seems to be no correlation between the expertise ratio and semantic originality. Additionally, and perhaps contrarily to intuition, new semantic associations (lower $r^{\mathcal{S}}$) do not correspond more to original teams (lower r^S) than to repeated teams — in other words, conceptual originality does not seem to be related to an original social composition of the underlying team.

Extending the notion of preferential attachment to socio-semantic hypergraphs enables the description of team assembly processes. This relies on a random baseline consisting of a null-model of evolving hypergraph. It features synthetic hyperedges which conserve the same number of agents and concepts as empirically observed, but arranged in an arbitrary manner (for a very recent theoretical overview, see Chodrow, 2020). In other words, randomly arranged socio-semantic teams $\widetilde{\Delta X_t} \in \mathscr{P}(S \cup \mathcal{S})$ are added at each period while respecting social and semantic size distributions of the empirically observed increment ΔX_t. Then, the discrepancy between empirical vs. simulated hyperedges for some property p yields an estimate of preferential team formation or preferential hypergraphic attachment: $\tilde{\pi}_t(p) = \left|x \in \Delta X_t \wedge p(x)\right| / \left|x \in \widetilde{\Delta X_t} \wedge p(x)\right|$. For instance, p may be chosen as some value of the expertise ratio. In this case, the results show a strong socio-semantic preference for groups made of a high or low proportion of experts (ξ close to 0 or 1 i.e., teams where none or almost all members just started working on a given concept for the first time; see Fig. 9a). We further observe a significant bias towards social and semantic repetition, consistent with similar results using flat measures on social graphs (Guimera et al., 2005): as may be seen on Fig. 9b, there is a high likelihood to repeat previous collaborations patterns (high propensity for high r^S), while the hypergraphic arrangement of concepts by a given team depends largely on the repetition of previous associations (increasing propensity with respect to $r^{\mathcal{S}}$).

3.3.2 Collectives at the meso-macro level: clusters

Going a little bit further up from the meso- to the macro- level brings us to the issue of aggregates, variously denoted as "communities" or "clusters", depending on the main discipline and aim of the respective studies. Contrarily to hypergraphs, there is here a long and now huge history of research, both in mathematical sociology and SNA, and in complex network and computational social science (established reviews in the respective fields may be found in Wasserman and Faust, 1994; Fortunato, 2010). In all generality, the former has rather had a tradition of using explicit algebraic patterns — starting with cliques (Luce, 1950) or structural equivalence (Lorrain and White, 1971) i.e., complete monopartite or bipartite subgraphs — while the latter rather has usually relied on looser statistical or algorithmic patterns, also as a result of dealing with larger graphs (Moody, 2001). In any event, the current notion of structural social aggregates, which I will conventionally denote below as “clusters”, may find its root indistinctly in the former or latter fields: be it a group of nodes such that “its members should have many relations with each other and few with non-members” (Alba, 1973) or “within which connections are dense, but between which connections are sparser” (Newman, 2004).

As said before, social clusters are also deemed to correspond, at least implicitly, to underlying semantic boundaries: a good clustering would for instance successfully differentiate oncologists from embryologists in academic collaboration networks; or workmates from schoolmates, in friendship networks. From a broader viewpoint, determining the success of cluster detection from relational data may generally be roughly assimilated to a socio-semantic mapping operation — even within a strictly social framework, clusters supposedly represent social circles with some underlying meaning, and this constitutes a frequent validation criterion in the literature. Here again, studies focusing on scientific dynamics have paved the way, by co-mapping individuals or journals and fields of knowledge (see e.g., Small and Griffith, 1974, Kreuzman, 2001 or Rosvall and Bergstrom, 2010, among many others).

Beyond that, research issues explicitly related to socio-semantic clusters may be categorized along two main domains. First, appraising the socio-semantic cohesiveness of clusters: to what extent are social clusters semantically cohesive or, vice versa, how semantic clusters are shared across groups of actors? Second, detecting intrinsically socio-semantic clusters: to what extent the joint social and semantic cohesiveness may be used to improve the characterization of clusters?

Socio-semantic cohesiveness. A sizable literature, already massive at the beginning of the 2000s, aims at coloring social clusters using semantic attributes (Brandes et al., 2001) or clusters (Börner et al., 2003). More precisely, topical clusters extracted from the semantic or socsem network are in turn projected on social networks. These clusters are either purely semantic or affiliation-based clusters i.e., they are based on either $\mathscr{s}$ or χ. Science (Vilhena et al., 2014; Raimbault et al., 2016) and online communities (Etling et al., 2010; Gaumont et al., 2018) are again typical playgrounds for this type of endeavors. The resulting semantically-colored social maps are oftentimes commented qualitatively and visually, in that such and such region is deemed typical of such and such community. Another stream of studies proceeds roughly the other way around by depicting semantic networks with respect to actors (Carley, 1994; Tancoigne et al., 2014; Basov et al., 2019): the goal is principally to exhibit common meaning structures across actors by focusing on representations and concept associations that are shared by pairs or groups of actors. Finally, joint socio-semantic regions may be identified on hybrid actor-concept networks gathering both node types ($S \cup \mathcal{S}$), either by using χ alone (Cambrosio et al., 2004; Callon, 2006; Hellsten and Leydesdorff, 2020) or, more recently, by relying on the full socio-semantic network i.e., by using additionally s and $\mathscr{s}$ (Roth and Basov, 2020): by contrast with the above-mentioned depiction of social clusters on semantic maps, or vice versa, cohesive regions are this time directly hybrid.

Beyond visual inspection, a whole set of studies has aimed at developing measures of the semantic cohesiveness of social clusters, in a variety of contexts: using links within (Davison, 2000) or between (Flake et al., 2002) web pages, blog networks Adamic and Glance (2005); Elgesem et al. (2015); Kaiser and Puschmann (2017), micro-blogs (Java et al., 2007; Himelboim et al., 2013). This strand has shown for instance how semantic cohesiveness changes as a function of link types or topics (e.g., Conover et al., 2011; Barberá et al., 2015; Garimella et al., 2016, see also Section 1.3 for more references), all of which reminesces about White (1992)'s networks of meanings. Ding (2011) interestingly raises the broader issue of a possible continuum in the alignment, or lack thereof, between social and semantic communities. Social clusters may cover a variety of semantic clusters, semantic clusters may spread over many social clusters. By challenging the assumption of widespread homophily in socio-semantic networks, we may further ask the question of the meta-structure of such systems in terms of more or less aligned social & semantic clusters: how to describe to what extent both types of clusters are aligned, under which conditions are they?

Socio-semantic clusters. Notwithstanding, another sizable portion of the literature directly assumes, rather than appraises, socio-semantic cohesiveness, and aims at building upon this premise to discover socio-semantic clusters per se. This goal may in turn be understood in two main different ways: either discovering bipartite clusters directly (principally using the socsem network χ), or combining social structure and semantics to uncover better or more precise social clusters (principally using the social and socsem network, s and χ).

The notion of bipartite clusters, or co-clusters, as cohesive groups of nodes of $S \cup \mathcal{S}$ and links of χ may for instance be illustrated by Mohr (1994) and Mohr and Duquenne (1997), which both focus on social assistance to define categories of actors (soldiers, mothers, blind people, etc.) who are provided similar categories of social relief. The former paper looks for co-clusters of actor types and relief concepts using a correlation-based approach normally applied to one-mode networks (Breiger et al., 1975), while the latter defines them more strictly as bicliques. Bicliques are maximal joint groups of actors and concepts connected in χ: this direct bipartite equivalent of cliques allows a lattice-based hierarchical analysis that I will detail further in Section 3.4. Note also here the interesting extension to tri-cliques by Jäschke et al. (2008) based on tri-partite relationships connecting agents, concepts (attributes) and artifacts (digital items) in so-called *folksonomies*.

One-mode cluster detection techniques have also been applied to two-mode networks, notably by proposing two-mode blockmodels (Doreian et al., 2004), or by formulating the monopartite measure of modularity (Newman, 2006) for bipartite graphs (Barber, 2007; Guimerà et al., 2007), or even by introducing the notion of two-mode cores (Cerinšek and Batagelj, 2015). They have also been applied to hybrid social and socsem networks: Zhou et al. (2009) discover clusters using random walks that may indifferently involve social and socsem links, while Yang et al. (2013) use a generative model of latent membership of nodes to communities which influences social links while taking into account attribute similarity i.e., socsem links. This brings us to the other class of methods that relies in all generality on the notion of latent categorical variables, and which is at the core of two very popular modeling techniques briefly evoked in Section 1.3: LDA and SBM. Broadly, these are generative Bayesian probabilistic models in the sense that they assume that latent variables, categories or blocks explain the generation of the observed links (recalling that actor attributes may be seen as links of χ).

The initial formulation of LDA (Blei et al., 2003) presupposes a latent, purely semantic layer of "topics" defined as distributions on words which explains the observed distribution of words in documents (see Blei, 2014, for an extensive review). The output may thus be seen as a weighted relation between documents and topics and, in turn, a weighted relation between topics and words. There is no other notion of cluster yet. The extension of LDA by Rosen-Zvi et al. (2004) called "Author-Topic" model soon introduced actors, more precisely authors behind documents, in that topics are distributed over authors, who then generate documents according to these distributions of topics, which in turn are still distributions on words. There is no social layer yet i.e., no social links. This appears in the "Author-Recipient-Topic" model (McCallum et al., 2005), whereby topic distributions depend on links between actors, or in the so-called "Topic Modeling with Network structure" (Mei et al., 2008), where topic distributions are improved using extra information from the social network. At this stage however, there is still no notion of social cluster: this appeared in parallel in Zhou et al. (2006c) who generate actor-concept clusters, either as sets of users who are associated with a topic set, or as sets of topics which are distributed over users. The simultaneous detection of topics (made of words) and communities (made of actors) happens in Chang and Blei (2009) and Liu et al. (2009), while the really symmetric detection of actor-concept clusters is achieved by Pathak et al. (2008) with the "Community-Author-Recipient-Topic" model adapted to sender-recipient (directed) social links, and by Sachan et al. (2012) with the "Topic User Community Model" for generic social networks. Both types of models jointly attribute, or discover, membership probabilities of users and topics to some latent communities i.e., to intrinsically socio-semantic clusters.

SBM, on the other hand, encapsulates the notion that social blocks possess some underlying semantic meaning: blocks are latent semantic parameters, even when only applied on social network data (Peixoto, 2018). When applied to affiliation data or socsem networks, SBM reveals a correspondance between actor blocks and concept blocks (Larremore et al., 2014; Gerlach et al., 2018). The aims of LDA and SBM appear to be merged in the "Stochastic Topic Block Model" (STBM, Bouveyron et al., 2018) which jointly attributes actors to blocks (and their inter-block connection probabilities) and to latent topics (and their word distributions).

3.4 Macro-level socio-semantics: phylogenies & lattices

Irrespective of the approach adopted to define them, socio-semantic clusters are actually socio-semantic hyperedges, though at a higher scale — they may nonetheless be expressed as elements of X. This further raises the question of their arrangement at a macro-level, either as a collection of hyperedges partitioning $S \cup \mathcal{S}$ or as a partially overlapping cover of $S \cup \mathcal{S}$. The literature already extensively addressed this issue in the case of social network clusters (and more broadly on networks featuring a single type of nodes), in terms of nested hierarchy (Moody and White, 2003), size distribution (Arenas et al., 2004) and general configuration (Rosvall and Bergstrom, 2008; Ahn et al., 2010); as well as for overlapping (Xie et al., 2013) or multilayer (Kivelä et al., 2014, §4.5) clusters.

Socio-semantic hierarchy. Beyond the above-mentioned question of a possible alignment between social and semantic clusters (Ding, 2011; Gerlach et al., 2018), the macro-level arrangement of socio-semantic clusters has received more attention only recently: in the CSS literature, it appears to be principally focused on detection efficiency (Melamed, 2014; Larremore et al., 2014) more often than the global structure and hierarchy per se. The configuration of bicliques in a given socsem network constitutes a notable and relatively old exception (Boeck and Rosenberg, 1988; Freeman and White, 1993), which is furthermore at the root of a series of contributions of mine focused on the overall organization (Roth and Bourgine, 2005), dynamics (Roth, 2006b; Roth and Bourgine, 2006; Roth, 2010) and also reducibility (Roth et al., 2006; Kuznetsov et al., 2007; Roth et al., 2008a; Klimushkin et al., 2010) of such socio-semantic clusters (for a review, see Roth, 2017). These studies built upon the notion of epistemic community (EC, Haas, 1992), which refers *a minima* to groups of actors sharing the same concepts and epistemic goals: solving a given socio-technical problem, advancing science in a given field, etc. ECs may be formalized as a dual set of actors altogether using the same concepts, or a biclique of χ — in set formulation, it is a maximal set of nodes $C \subseteq \mathscr{P}(S \cup \mathcal{S})$ such that $\forall(a, c) \in (C \cap S) \times (C \cap \mathcal{S})$, $(a, c) \in \chi$ and there exists no superset of C where the same property holds.

As subsets, hyperedges induce a natural hierarchy based on set inclusion which is best represented as a lattice. Socio-semantic hyperedges feature a natural dual hierarchy: an EC C_1 can be said to be more general than another EC C_2 when the actor set of C_1 contains that of C_2 and thus, dually, the concept set of C_1 is included in that of C_2. This dual relationship defines a partial order $\geq_{\text{ec}}$ on bicliques such that for two ECs C_1 and C_2,

$$C_1 \geq_{\text{ec}} C_2 \iff (C_1 \cap S) \supseteq (C_2 \cap S) \iff (C_1 \cap \mathcal{S}) \subseteq (C_2 \cap \mathcal{S})$$

In turn, partially-ordered ECs may be arranged in a lattice, and more precisely a particular instance of a Galois lattice (Barbut and Monjardet, 1970; Freeman and White, 1993), which is also the core focus of the "Formal Concept Analysis"

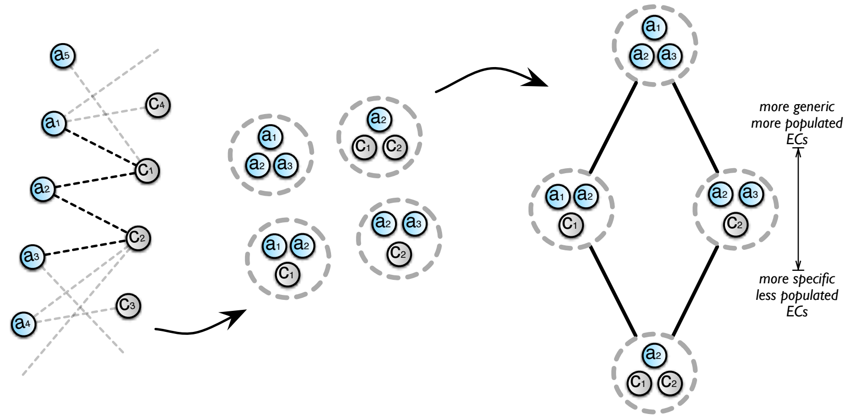


Figure 10: Illustration of the construction of a socio-semantic lattice $\mathscr{L}$, from left to right: (i) a bipartite graph features usage of concepts c_1, c_2, etc. by agents a_1, a_2, a_3 (part of χ); (ii) maximal groups of agents using the same concepts are then extracted (they are meso-level socio-semantic hyperedges and thus form a type of socio-semantic hypergraph X) and finally arranged into a hierarchical socio-technical lattice $\mathscr{L}$ (being a high-level socio-semantic hypergraph with the partial order $\geq_{\text{ec}}$). Note that S and $\mathcal{S}$ are by definition bicliques and respectively upper and lower bounds of all ECs.

(FCA) community (Ganter and Wille, 1999) where bicliques are called "formal concepts" and where, generally speaking, S is a set of objects while $\mathcal{S}$ is a set of attributes. The socio-semantic lattice $\mathscr{L}$ is eventually based on $S \cup \mathcal{S}$, the order relation $\geq_{\text{ec}}$, and the set of all socio-semantic bicliques/ECs of X. See a toy illustration on Fig. 10, vertically represented as a Hasse diagram (Davey and Priestley, 2002). Navigating $\mathscr{L}$ from top to bottom is equivalent to exploring socio-semantic communities from the most generic to the most specific ones. Moreover, since ECs may have more than one parent and more than one descendant, they are well-suited to the representation of non-Aristotelian taxonomies, allowing for the membership of an item to several categories.

An empirical illustration from Roth (2010) is given on Fig. 11. It relies on MEDLINE bibliographical records mentioning the word "zebrafish" between 1990 and 2003, in order to capture the emergence, consolidation and institutionalization of the scientific community interested in this model animal. This yields a socio-semantic network χ whose social boundaries are semantically defined i.e.,

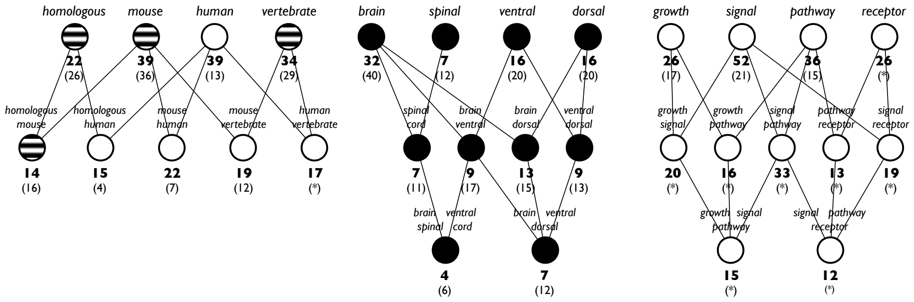


Figure 11: Evolution of the socio-semantic macro-structure of the zebrafish community from the period 1990-95 to 1998-2003, using pruned socio-semantic lattices focused on the top (most generic and populated) ECs. Each node is an EC labeled with its concept set, and its actor set size as a percentage of the total population. Figures are in bold for 1998-2003 and in italics and inside brackets for 1990-1995, showing the lattice evolution over the period. White and black ECs experienced a marked population increase or decrease (more than 15%), dashed ECs stagnated (less than 15%), stars denote new ECs in 1998-2003. *From* Roth *(2010).*

both extending to and limited to this field whose members mention "zebrafish" almost systematically in the abstract. The data is divided into three slightly overlapping 5-year periods, focusing on a fixed number of concepts (70) among the most significant across all periods, and on about 1,000-10,000 actors from the first to the last period. These figures are way enough to generate an intractably huge number of bicliques (Ganter and Wille, 1999) and, for comparison purposes, require to concentrate on random samples of a fixed size of 250 actors for each period. Even so, lattices are huge: the first period lattice contains a couple hundred thousands of bicliques, which calls for further pruning heuristics (see below). To start with, we selected the 20 top-ranking ECs for each lattice based on a simple score combining population size (in terms of actors, to favor communities gathering a sizable portion of the field) and distance to the lattice top (to favor more general communities). The longitudinal comparison of socio-semantic lattices describes the high-level evolution of the social distribution of cognitive tasks within the field. In particular, three main research areas may be identified, organized around three subsets of concepts and corresponding actors: (i) the study of biochemical signaling mechanisms, involving pathways and receptors; (ii) comparative studies focusing on similarities and differences between humans, mice, zebrafish as vertebrates; (iii) the examination of the nervous system and brain development. The first and, to a lesser extent, the second subfields grew in importance within the community at the expense of the last field: research on brain and spinal cord decreased and its relationship with ventral and dorsal aspects became weaker. On the other hand, the community started to venture into signaling issues; which is partly explained by the emergence of a more general background trend in molecular biology.

The computational explosion linked to the representation of lattices and, more broadly, socio-semantic hyperedges, even for small socio-semantic systems and datasets, poses both a quantitative problem, in terms of calculation or visual representation, and a qualitative problem, as regards the relevance of taxonomies based on a tremendous number of categories. The above-mentioned selection method focused on the top of the lattice, following the then current state-of-the-art based on so-called "iceberg" lattices (Stumme et al., 2002) which came with a significant price in terms of resolution: for instance, it typically disregards mid-sized and niche ECs. In this regard, I could contribute to improve systematic lattice reduction methods (Roth et al., 2006; Kuznetsov et al., 2007; Roth et al., 2008a) by operationalizing the notion of *stability* (Kuznetsov, 1990), further enriched in Klimushkin et al. (2010) with the notions of concept "probability" and "selection". More precisely, a formal concept, or socio-semantic hyperedge, is "stable" if the absence of some of its items or properties does not prevent its existence as a pattern. Instead of removing the bottom part of the lattice using a relatively arbitrary size threshold, our combinatorial approach leverages the tendency of the lattice to be self-similar and to contain many duplicate patterns. It prunes slightly redundant bicliques *everywhere*, at any size, from top to bottom, and has been widely used in the FCA community (Poelmans et al., 2013).

Socio-semantic phylogenies. This example illustrates two main issues of the macroscopic description of intrinsically socio-semantic structures: pattern selection to build meaningful (static) taxonomies, and inter-temporal matching to build meaningful (diachronic) phylogenies. They are both, to some extent, a selection problem, which is already well-known in the case of social cluster detection, both statically (as a resolution limit, see Fortunato and Barthelemy, 2007; Traag et al., 2011) and diachronically (especially community instability, see Rossetti and Cazabet, 2018, for a review). The diachronic case is particularly interesting with respect to the above example as it also raises the issue of pattern redundancy and stability, however in a temporal perspective: communities are partly redundant, partly stable across time.

Typically, inter-temporal correspondence may be assessed longitudinally, either by using network snapshots (clusters at t are associated with clusters with similar members at t', see Doreian, 1979; Hopcroft et al., 2004, for instance) or link recurrence (the stability of links observed between t and t' defines the group, as in Palla et al., 2007), thereby assuming that social entities only exist by way of their temporal stability (Abbott, 1995). The identification of semantic clusters is naturally isomorphic and its longitudinal formulation is also the object of relatively recent research, especially the automatic construction of temporal knowledge maps (Rosvall and Bergstrom, 2010; Shahaf et al., 2012) and phylogenies (Lancichinetti and Fortunato, 2012; Chavalarias and Cointet, 2013). In these fields, I could contribute to longitudinal semantic cluster analysis using network snapshots (representing the evolution of semantic clusters across time in a unified multi-scale representation Chavalarias et al., 2011).

Inter-temporal correspondence may also be assessed dynamically, based on what may loosely be called temporal networks and which preserves the fact that the primary empirical material is made of links that appear across time rather than aggregate network snapshots; thus mainly looking for inter-temporal clusters of links. An early dynamic social cluster analysis techniques consists in introducing links connecting the same nodes across various periods of time (Ben Jdidia et al., 2007; Mucha et al., 2010). Here, I could also propose a contribution focused on edges rather than snapshots in (Mitra et al., 2012). It aimed at detecting clusters in a meta-graph allowing inter-temporal connections between distinct nodes: typically in citation networks (science or blog posts) where edge extremities connect nodes at different times, by definition rather than by construction. A more recent approach features so-called stream graphs and link streams (Viard et al., 2016) where links are intrinsically inter-temporal: they have a temporal thickness, which appears to pave the way to a truly dynamic viewpoint on network evolution (Latapy et al., 2018).

By definition, the interactional analysis of social clusters steers clear of intensional properties: in a dynamic perspective, this means that the old sociological question of the perpetuation of social groups[2] is appraised through the stability of interactional structures across time rather than the persistence of their attributes. In the case of knowledge community mapping, socio-semantic hypergraphs could here again be part of the solution toward developing a unified formalism to describe the dynamics of both relational and topical clusters (expressed in S and $\mathcal{S}$ based on s and $\mathfrak{s}$), and of joint socio-semantic clusters (expressed in X based on χ or a mixture of s, $\mathfrak{s}$ and χ). Beyond this, the characterization of phylogenies of intrinsically socio-semantic structures remains a challenging and open field.

[2] "*The most general case in which the persistence of the group presents itself as a problem occurs in the fact that, in spite of the departure and the change of members, the group remains identical. We say that it is the same state, the same association, the same army, which now exists that existed so and so many decades or centuries ago. This, although no single member of the original organization remains.*" (Simmel, 1898, p. 667)

4 Socio-semantic issues

Understanding morphogenetic dynamics is admittedly a preliminary to the study of more sophisticated social cognition processes happening within socio-semantic systems. Let us now review several contemporary issues which derive from, or develop upon, the evolution of the socio-semantic structure: phenomena of content diffusion, the related question of influence and authority, the co-evolutionary emergence of fragmentation and polarization, and the role of algorithmic devices in socio-technical systems.

4.1 Diffusion processes

Knowledge diffusion, in the broad sense, is one of the most salient social cognition phenomena, whereby a socio-semantic system collectively interprets, adopts, transmits some unit of information, in a both collective and self-organized manner. Socio-semantic networks, especially s and χ, are naturally adapted to the joint description of the potential social paths upon which information travels or may travel, and of the attribution of some belief, opinion, interest, or knowledge to actors as socsem links. This framework does not entail any decision as to whether information propagates on a relatively fixed structure (χ depends on s), or social connections themselves are influenced by the distribution of information (s depends on χ) — let alone the possibility that some content propagates better (χ depends on $\mathfrak{s}$, as in Godart and Galunic, 2019). This agnosticism is assuredly compatible with the perennial chicken-and-egg debate regarding peer selection and influence, and more precisely the difficulty of assuming whether social ties postdate or predate cultural transmission (Vaisey and Lizardo, 2010). Despite some skepticism about the possibility of appraising them empirically from observational data (Shalizi and Thomas, 2011), many studies have nonetheless tried to appraise selection and influence jointly (Snijders et al., 2007; Crandall et al., 2008; Aral et al., 2009; Lewis et al., 2012; Christakis and Fowler, 2013).

The literature oftentimes did make a decision, however, on the primacy of one realm over the other. As discussed in Section 3.2, many studies focus on the dynamics of s as a dependent variable on χ. There is, on the other hand, a large body of research that is devoted to the dynamics of χ while assuming a fixed social network structure s. This second perspective is also very much compatible with an epidemiologic viewpoint. It thus attracted a lot of natural science endeavors translating classical (biological) epidemiological models into cultural epidemiological models (albeit not in the sense of Sperber, 1996).

From this viewpoint, knowledge diffusion is also a socio-semantic morphogenesis issue, even though the eventual focus lies on macroscopic patterns i.e., it principally aims at articulating micro-level processes on χ, dependent on s, with macro-level observables on χ, such as the extent or speed of propagation in the whole system. This issue had been initially addressed by social scientists relying on ethnographic studies. They exhibited determinants of knowledge transmission and adoption behaviors (Robertson, 1967; Rogers, 1976; Granovetter, 1978; Burt, 1987; Valente, 1996) or proposed normative models based on stylized hypotheses derived from established theoretical frameworks (Ellison and Fudenberg, 1995; Abrahamson and Rosenkopf, 1997; Deroian, 2002). The importance of correctly designing and understanding the underlying interaction structures grew across time. For a while, this still happened on normative rather than descriptive grounds, i.e. with stylized networks (Goyal, 2005; Morris, 2000; Cowan and Jonard, 2004). An early call by Rogers (1976) to use empirical networks long remained a distant aim, especially in the absence of relevant data.

A stronger empirical approach was finally encouraged by the recurrent observation of the peculiar structure of large-scale interaction networks — notably their heterogeneous connectivity structure, the existence of clusters and of a small network diameter. In the wake of a key result by Pastor-Satorras and Vespignani (2001) suggesting that such networks have radically distinct epidemiologic properties from those of uniformly random networks, the literature on diffusion models started to take network structure into account to specifically study the contrasted effects of various topological assumptions. Authors started to investigate diffusion phenomena on networks by building upon the epidemiological literature and its various SI-models (such as SI, SIS, SIR, SIRS, etc., where S stands for Susceptible, I for infected, and R for recovered, and actors transition between these states) with a key difference: the introduction of iterative processes with networks at their core (Lloyd and May, 2001; Eguiluz and Klemm, 2002; Newman, 2002), thereby strongly diverging from macro-level approaches based on differential equations (Hethcote, 2000).

A variety of diffusion models and processes have then been translated and tested in an eminently networked framework (Amblard and Deffuant, 2004; Ganesh et al., 2005; Crépey et al., 2006). Later studies progressively addressed communication science topics, discussing roles such as “influencers” (Watts and Dodds, 2007;

Kitsak et al., 2010) and involving more sophisticated contagion processes diverging from binary infections, accompanied by validation on real-world data. Over the last two decades, the community has produced a wealth of empirical studies connecting (micro- and macro-level) topology, (micro-level) behavior and (macro-level) diffusion patterns, around questions including:

- *Which empirical contexts?* Looking concretely at the evolution of scientific topics across collaboration links (Zhou et al., 2006b), the propagation of simple digital artefacts such as favorite pictures in photo-sharing platforms like Flickr (Cha et al., 2009a), URLs in blogspace (Cha et al., 2009b) or hashtags in Twitter (Weng et al., 2012); or more complicated content such as user-created gestures in virtual worlds (Bakshy et al., 2009) or rumors on Facebook (Friggeri et al., 2014).
- *Which topological features?* Discussing the role of either (i) ego-centered properties such as connectivity (e.g., influencers may, or may not, be hubs Cha et al., 2010; Borge-Holthoefer et al., 2012), the importance of betweenness centrality and so-called "weak links" (Grabowicz et al., 2012) or the joint effect of both position and connectivity (Gonzalez-Bailon et al., 2011; Ma et al., 2013), (ii) meso-level topology, by insisting rather on the importance of clusters rather than individual properties (Watts and Dodds, 2007; Bakshy et al., 2011) and the uneven propagation across them (Weng et al., 2013; Bruns and Sauter, 2015), or (iii) macro-level influence flows between communities (Chikhaoui et al., 2017).
- *Which adoption behavior?* Refining the knowledge of micro-level contagion, generalizing the classical threshold (Granovetter, 1978; Valente, 1996) and (iterative) cascade models (Leskovec et al., 2007), or introducing so-called "complex contagion" based on multiple exposures (Centola and Macy, 2007; Romero et al., 2011).
- *Which diffusion observables?* Characterizing for instance the shape of propagation trees, or cascades (McGlohon et al., 2007; Goel et al., 2016), the extent of diffusion, its speed (see e.g., Iribarren and Moro, 2011, for a combined study of both), or its prediction (Cheng et al., 2014).
- *Which dependence on content?* Describing more broadly the existence of socio-semantic correlations in the diffusion process, focusing in particular on the fact that some information may be more relevant to some actors than others (Wu et al., 2004), spread differently depending on intrinsic properties (Kuhn et al., 2014; Vosoughi et al., 2018) or may be more contagious in similarly-minded clusters (Bakshy et al., 2015).

Influence of semantic properties on diffusion. Most of my work in this area contributes diversely to each of these questions, with a marked attention to the measurement of empirical processes, and the importance of taking into account the joint role of network structure and content dynamics. Toward the second half of the 2000s, many formal diffusion models still relied on (i) stylized networks — either uniformly random (Erdős and Rényi, 1959) or preserving empirical connectivity distributions (using e.g., a configuration model, Molloy and Reed, 1995), (ii) stylized inter-individual processes — often cascade or threshold models (Kempe et al., 2003). It was unclear how likely these assumptions could accurately render real-world phenomena. To check this, I first contributed to develop a concise benchmarking framework based on simulated diffusion models, while introducing a slight (though more realistic) variation with respect to classical assumptions (Cointet and Roth, 2007a,b). Using either real-world interaction structures or empirically-measured adoption processes yielded significantly distinct outputs, calling for more precision in appraising empirical phenomena before drawing conclusions on knowledge transmission.

I then endeavored at showing that empirical propagation dynamics were heterogeneous with respect to the semantic type and the structural position of nodes. In a first study, we used hidden Markov chain models to exhibit systematic precedence phenomena between groups of sources, using a dataset of websites including online news media and politically-oriented blog sites (Cointet et al., 2007). We could thus demonstrate, at a macro level, the existence of complex intertemporal relationships between node affiliation/type and topic occurrence — significantly more complex than could appear in the studies discussing precedence issues between news and blogs (Lloyd et al., 2006). This issue was further detailed by integrating the topology in a subsequent work which still focused on political blogs, yet articulated precedence relationships and network structure. More precisely, we studied in Cointet and Roth (2009) the propagation of hypertext links as a function of local neighborhood properties. At a local level, we showed the influence of past connectivity on influence (in a nutshell, bloggers who received more attention in the past will have more influence), yet in a non-linear fashion: the association is much milder for nodes having garnered a markedly weak or stong attention. At a meso-level, we confirmed that weak links connecting "distant" network areas were preferred diffusion pathways (as expected, see Onnela et al., 2007), but only up to a certain point, after which influence sharply decreased: in other words, the way weak links facilitate diffusion was bounded.

Finally, I recently started developing a research direction that takes into account the semantic structure: on the one hand, at the very micro-level of reformulation processes; on the other hand, at the more macro-level of the navigation of actors on a content network. First, as exposed in section 1.1, Sperber (1996)'s cultural epidemiology postulates cognitive attractors to explain cultural representation similarity and stability at the social level, despite extremely noisy individual information reproduction mechanisms, even in controlled environments (Moussaïd et al., 2015). Elementary forms of cognitive bias could be exposed *in vivo* in Lerique and Roth (2018), using data on mutations in quotations in a large blog post corpus (Leskovec et al., 2009). We could identify the factors influencing man-made and likely involuntary substitutions in reproducing quotations, by drawing on standard linguistic variables, such as the age of acquisition of a word, its frequency, its length and the gap between the incidence of the original term and that of the substitute term. Second, I examined how the structure of content networks affects information accessibility. This issue is more broadly related to the articulation between topological and semantic confinement, which is at the core of current studies on online echo chambers, whereby potential navigation paths likely affect the access to cross-cutting content (Bakshy et al., 2015; Garimella et al., 2018; Cinelli et al., 2021; Morales et al., 2021). Focusing on the topological navigation landscape defined by non-personalized YouTube recommendations, Roth et al. (2020) demonstrate that ego-centric content graphs with higher entropies (in terms of visited nodes) counter-intuitively exhibit lower diversity (in terms of the distinct number of accessible videos). Put differently, in this case, while a user may appear to visit more locations, this however occurs within an overall smaller space: some videos are at the root of an isotropic navigation in a more limited space of videos. These videos are additionally amongst the most viewed, which sheds light on how recommendation devices contribute to concentrate user navigation (see below Section 4.4).

4.2 Authority: structural positions and temporal patterns

While diffusion processes occur at an admittedly short-scale, studying influence on the longer term relates to a more crystallized and temporally aggregated notion often denoted as "authority". In other words, authority may be seen as influence over a higher temporal scale. One of the simplest ways to appraise authority within a network framework consists in looking at the configuration of accumulated references (incoming links) over a certain period of time, and connecting it with short-term phenomena.

The above example of Cointet and Roth (2009) where we framed influence with respect to past attention could be considered as a preliminary attempt in this regard. I pursued this path in two additional directions, attempting again at connecting content and structure. First, by focusing specifically on temporal phenomena, authority, and topics: in Menezes et al. (2010a,b), we linked the precocity of actors in evoking some issue (appearance of a socsem link to the issue) with their accumulated structural centrality in the social network as a measure of authority. Using a corpus of political blogs, we detected nodes who are "systematically" prompt (in probability) to address some topics i.e., who are early relative to the rest of the system, at short time-scales. This relied on a hybrid approach combining text mining, signal analysis and peak detection. We could then compare this precocity to positions in the underlying network, positions being properties defined on a longer time-scale, thereby connecting the literature on attention peaks (e.g., Crane and Sornette, 2008; Lehmann et al., 2012) and influence measures (e.g., Agarwal et al., 2008). This led to a typology of actors under the form of a double-dichotomy based on structural authority and temporal precedence, accurately distinguishing *copycats* (early birds with low authority) from political figures (with higher authority: either emerging and early, or established and late).

Second, by proposing a topological model of the configuration of topical blog communities (Cardon et al., 2011, 2014), later extended to Twitter (Roth and Hellsten, to appear) and which could be easily be adapted to other contexts such as scientific communities. In the initial study of 2011, we distinguished endogenous and exogenous authority, respectively from inside or outside of ego's topical community (topical tagging had been done manually by a team of librarians based at one of our partners: e.g., sport, cooking, politics, etc.). We combined it with endogenous and exogenous activity to provide a 4x4 matrix model describing a variety of relevant structural positions occupied by nodes in their own semantic

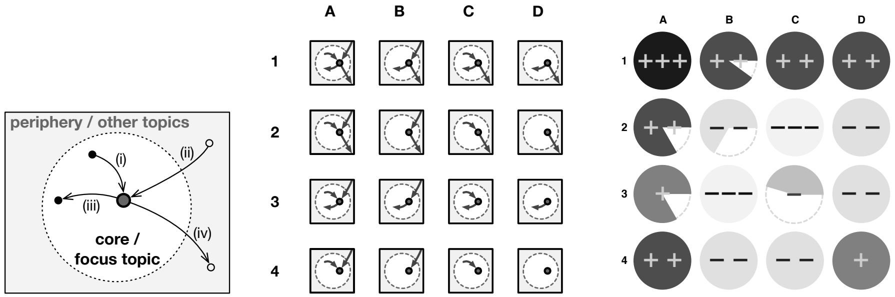


Figure 12: Socio-semantic, or structural-topical model introduced in Cardon et al. (2011, 2014). Left: It relies on a double dichotomy between incoming or outgoing links, internal (white area) or external (gray area) to a topical territory or a core. Middle: From there, 16 cells are defined, depending on whether a node is most connected (e.g., top 20%), or not, in each of the 2x2 categories. Right: It has been applied on French and German blogspaces, for 6 (x2) topical territories (cooking, IT, politics, etc.). Symbols correspond to over- or under-representation (+ or -), weak or strong (+++ vs. +), and are here averaged for all territories (full disks indicate that all territories behave similarly). It exhibits remarkable structural positions which are always over- or under-represented, while some slots, such as C3, exhibit a mixed pattern: they exist only for self-centered topics (home cooking, worn fashion) which are thus truly community-centric territories (as in Rheingold, 1993), not for public space topics (politics, IT progresses).

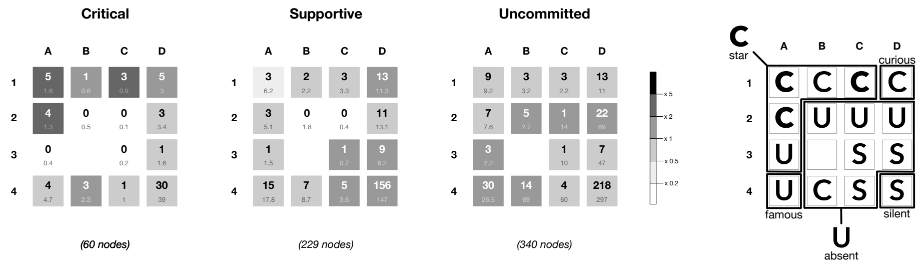


Figure 13: Application of the model of Fig. 12 to a Twitter subspace focused on users discussing the 5th Assessment Report (AR5) of the Inter-Governmental Panel on Climate Change (IPCC). The core is now defined over the most active users, who were manually labeled as supportive, critical, or neutral toward climate change science (Roth and Hellsten, to appear). The three matrices on the left describe how each alignment is more or less present than expected if position did not matter i.e., if it were uniformly represented in each slot. The matrix on the right shows that critical discourse occupies the dominant positions, even if they are the minority.

territory. This socio-semantic, or structural-topical model exhibited regularities which are partly independent and partly dependent on underlying topics. It thus contributed both to topically refine the notion of online authority (Hindman et al., 2003) and to enrich blogspace typologies (Karpf, 2008). The model has been statistically extended and doubled with a qual-quant approach in a wide comparative study accross the French and German web in Cardon et al. (2014); see a succinct overview on Fig. 12. The very recent extension to Twitter (Roth and Hellsten, to appear) aims at appraising the intertwinement of authority and activity in the specific case of climate change discussions, where the conversation network features interactions which are not homophilic and occur between skeptics and supporters of the scientific basis of climate change (Pearce et al., 2014). The socio-semantic mixing of this cross-cutting social network blurs the structural boundaries between the opposing poles of a debate. In this situation, our model carries out a socio-semantic positional analysis of the active core of the conversation network and demonstrates that minority alignments (skeptics) occupy nonetheless the dominant authority and activity positions — see Fig. 13.

A complementary field of investigation consists in detailing the process of accumulation of authority itself, more specifically its temporal features and its relationships with quality (Roth et al., 2011b). In an online socio-technical system such as Wikipedia, where quality is a product of collaborative dynamics (Wilkinson and Huberman, 2007), I worked on the processes of establishment of reliable knowledge. I quantitatively described the various phases of addition of links to external references and the characteristics of their contributors (Chen and Roth, 2011), a small portion of which gathers most of the qualified edits. A later paper by other authors (Forte et al., 2018) furthered the exploration of this activity by introducing the concept of "information fortification", the addition of citation links in the specific context of controversies. These processes are contrasted with those arising in scientific communities. Roth et al. (2012b) also studied the dynamics of scientific citation networks by connecting future attention (measured by their citational impact i.e., incoming links after publication) to past attention (i.e. outgoing links at the time of publication). Put shortly, ego-centered past-oriented citation links of a paper inform about its ego-centered future-oriented citation links, thus its relevance for the community. Up to some rescaling, citation dynamics exhibit a universal behavior consistent across disciplines. This suggests that future citations may be predicted from the structure of references upon publication. Papers with above-average citation appear to focus extensively on their own recent subfield — as such, citation counts essentially select what may plausibly be considered as the most disciplinary and normal science — whereas papers which exhibit a peculiar dynamics, such as some resurgence after some time has passed, are comparatively poorly cited, despite their plausible durable relevance.

4.3 Co-evolutionary models of polarization

Modeling polarization, fragmentation, and, more broadly, the emergence of homogeneous socio-semantic clusters, is a notoriously socio-semantic morphogenesis issue, with a generative rather than descriptive viewpoint. It integrates a small number of micro-level hypotheses, especially in terms of social and socsem PA and adoption (Sections 2.1, 3.2, and 4.1), to reproduce a certain number of macro-level structural observables, in terms of what has been discussed in Sections 3.3.2 and, to some extent, 3.4. The decade-old review by Gross and Blasius (2008) dichotomized complex network research into two strands, focused either on the dynamics of network (morphogenesis per se) or on the dynamics on networks (e.g., diffusion). Both may be combined in a co-evolutionary framework where each interacts on the other: node states affect topological evolution which determines topology, which in turn affects local dynamics and thus states. At the time, most existing works were nonetheless principally focused on quite simple cognitive states closely related to game theory or SI models — and more specifically the adoption of a binary behavior in a population (i.e., A or not A), which generally narrowed the scope to bi-polarization.

There were already some exceptions. Two pioneer works from the end of the 1990s must first be mentioned here. They inaugurated a broad family of so-called "cultural dynamics" models. On the one hand, Gilbert (1997, briefly evoked in Section 3.1) introduced a co-evolutionary dynamics between some semantic space and some actor-like space. The latter space is made of scientific papers associated to some position in the former space, which is modeled as a two-dimensional discrete vector ('kenes'). The position of a new paper depends on some averaging of the positions of a subset of the neighboring existing papers that they cite: in other words, the generative process consists of the addition of social links (from citing to cited papers) and some form of socsem links (from papers to 2D vectors) when considering that each new paper induces the creation of a new semantic node. The model eventually demonstrates that such a simple dynamics suffices to produce socio-semantic clusters (groups of citing papers of similar semantic positions) and connectivity heterogeneity (some papers receive substantially more citations). On the other hand, Axelrod (1997) proposed an agent-based model where each agent possesses m cultural dimensions taking m' possible values. Note that this semantic space may be represented as $m \cdot m'$ semantic nodes divided into m classes, each agent being socsem-connected to exactly one node among each group. The social network is a regular rectangular grid where virtually every agent is connected to exactly four nodes. Each model step focuses on a random social link connecting two actors, who then interact with a probability proportional to their socsem-similarity; in which case one of the agents rewires one of their socsem link similarly to the other agent. The model demonstrates the emergence of culturally homogeneous clusters i.e., that micro-level convergence may lead to macro-level differentiation.

Several refinements have been proposed, such as the inclusion of some notion of memory (or stickiness of socio-semantic links) to explain the emergence of distinct norms (Axtell et al., 2001), the introduction of heterophobia (i.e., repulsion toward unalike agents, Macy et al., 2003) and of the possibility of removing social links in case of dissimilarity (Centola et al., 2007; Holme and Newman, 2007), or the existence of negative interactions (i.e., interactions that increase dyadic dissimilarity Flache and Macy, 2011; Smaldino et al., 2017) — all of which further illustrates the apparent paradox between local homophily and global differentiation.[3] These models are essentially based on dyadic interactions between agents, while socio-semantic dynamics also rely on group-level interaction patterns, as is the case in scientific teams (Section 3.3.1). In this regard, I could propose meso-level co-evolutionary dynamics based on homophily and some preference for repeated interactions (Roth, 2006a, 2008a). A similar modeling framework may be found in (Sun et al., 2013). This leads to the emergence of hierarchical socio-semantic clusters and functional differentiation in scientific communities which somewhat resonates with much earlier works such as Blau (1970).

These models also contrast in form, yet not in spirit, with so-called "opinion dynamics" models (Castellano et al., 2009), which are based on a single yet continuous cultural dimension. As such, they would not directly fit a socio-semantic network framework. However, they provide equally extremely stimulating explanations to the co-evolutionary emergence of socio-semantically segregated clusters. They originally feature interactions based on a similarity threshold (Deffuant et al., 2000), possibly compounded by the topology of an underlying social network (Am-

[3] This reminds the paradox raised in the formulation of cultural contagion, whereby imperfect micro-level idea transmission nonetheless leads to macro-level cultural homogeneity.

blard and Deffuant, 2004), and typically examine the role of such or such effect on the emergence of such or such number of clusters, in particular convergence or emergence of extreme factions. For an extensive recent review, see Flache et al. (2017). Worth remarking is that both cultural and opinion dynamics models have also been extended to take into account the role of macro-level effects, especially the contribution of a mean field to top-down agent synchronization (respectively González-Avella et al., 2007 and Hegselmann and Krause, 2015).

On the whole, these synthetic models raise concrete and contemporary questions on the existence of fragmentation and of possibly reinforcing socio-semantic clusters, often denoted as "echo chambers", in online public spaces. The early optimism (e.g., Rheingold, 1993) about the potential of online deliberative spaces to stimulate the rational convergence of opinions and the emergence of the ideal public sphere envisioned by Habermas had already suffered a variety of criticisms at the turn of the 2000s (Dahlberg, 2001). Specialized social groups with similar semantic interests have been increasingly seen as a threat to the exposure to cross-cutting content and diverging viewpoints (Van Alstyne and Brynjolfsson, 1996; Sunstein, 2007; Lev-On and Manin, 2009). In short, feedback loops induced by the "selection-influence" couple may have conflicting effects, especially as homophily, heterophobia and confirmation bias may reinforce prior beliefs and thus polarization. Social influence may reinforce similarity, yet social influence may reinforce divergence, as the above-mentioned modeling state of the art also shows: to use the trichotomy of Flache et al. (2017), these joint phenomena may be observed regardless of assuming assimilative social influence (influence increases similarity), similarity bias (the other way around: similarity influences interaction), or repulsive influence (usually combining assimilation with repulsion). Empirical studies exhibit the same tension (Barberá, 2020): some works indicate socio-semantic segregation, especially in affilation networks (citation, retweet, follower, subscription links), some works indicate the opposite, especially in interaction networks (conversation, mentions, quotes, etc.); sometimes a mix of both (Cinelli et al., 2021). Similarly, many recently-proposed models aim at reproducing the emergence of echo chambers under a variety of realistic assumptions, including mean-field influence or heterophobia (e.g., Wang et al., 2020; Sasahara et al., 2020), which is not unlike the massive effort over the 2000s to reconstruct power laws and clustering in social networks with a myriad of generative mechanisms — arbitrating between the various and sometimes discordant explanations may be perplexing. Given the current under-determination of theories by facts, it is nonetheless possible to hypothesize that various types of forces occur concurrently, albeit differently for distinct networks of meanings (interaction, affiliation, etc.) and for distinct topics — which likely calls for richer, multilayer socio-semantic models of online echo chambers.

4.4 Socio-technical systems and devices

The question of online polarization further emphasizes the intrication of socio-semantic dynamics with the devices that host them: each online platform configures a specific socio-technical setting with its own affordances and artificial ways of mediating access to content and to others. Using Twitter means posting short messages, following, mentioning, retweeting other users. Using Facebook means creating so-called "friendship" ties, publishing content on one's own personal page, commenting anywhere, tagging content with a variety of emotions (like, awe, anger, etc.). Using Reddit means contributing generally longer messages in reply to other messages in a tree-like fashion within discussion "threads" on some topic or subtopic, voting them up or down. Each of these systems defines a distinct type of online society whose socio-semantic dynamics are not necessarily easy to disentangle from the usage grammar induced by the platform and its algorithmic environment.

Consider Wikipedia and even wikis in general. From a technical viewpoint, they constitute a particularly simple example: users can essentially create and edit articles, discuss them in dedicated meta-pages, and, sometimes, meta-discuss the wiki rules. Put differently, there is "not much more" to study than the endogenous user and content dynamics happening within the bounds and technical possibilities of a rather straightforward wiki interface: typically, wikis do not feature any sort of algorithmic recommendation, which I will address below. Yet, complex *artificial* societies obviously have thriven: artificial, in the sense that interaction is constrained by the artificial rules of the underlying software. Wikipedia, as a lively and hierarchized community of hundreds of thousands of regular contributors, generated a wealth of qualitative and quantitative studies which may provide further insights on general social science phenomena — such as apprenticeship (Bryant et al., 2005), controversies (Brandes and Lerner, 2008) and democratic deliberation (Leskovec et al., 2010b), collaboration (Keegan et al., 2012), authority attribution (Chen and Roth, 2011; Forte et al., 2018) — yet are, above all, Wikipedia social science: the science of a certain society where the interface plays a certain role in disciplining both socialization and content creation.

Online communities have been an essential aspect of my research directions, not least for the possible in vivo observation of social cognition processes. Beyond the various explicitly socio-semantic studies evoked so far, I could also more anecdotally show how users evaluate content in ad hoc social networking platforms, establishing for instance a link between conformism in book ratings on the aNobii platform and the underlying network activity: users moderately diverging from the norm were more likely to have more "friends" or "neighbors" (Chen and Roth, 2012). Focusing on two types of content-based online communities, wikis (Roth, 2007d; Roth et al., 2008b,c) and photo-sharing platforms (Chen and Roth, 2010; Taraborelli and Roth, 2011), I could further show how their socio-semantic development (in a basic acceptation: population and content size) correlates with both governance and network features.

Algorithmic devices. The picture is further complicated by the potential interference of algorithmic recommendation (Salganik et al., 2006; Epstein and Robertson, 2015; Roth, 2019a). For one, the role of recommendation devices in fostering serendipity is assuredly at the heart of a fast-growing literature. Contrarily to clear-cut popular assumptions about so-called "filter bubbles", the jury is still out as to whether algorithms increase consumption diversity (Bakshy et al., 2015; Aiello and Barbieri, 2017; Möller et al., 2018; Puschmann, 2019) or not (Nikolov et al., 2015; Anderson et al., 2020; Roth et al., 2020). At least, explicit personalization or "self-selection" (whereby users voluntarily select or prioritize some sources e.g., by "liking" or subscribing to them) consistently appears to induce algorithmic reinforcement and confinement (Zuiderveen Borgesius et al., 2016). Nonetheless, most of these results apply at the aggregate level, without distinguishing subpopulations of users who may differently use or respond to algorithmic guidance. A few studies expressly differentiate users who are eager for recommendation (Nguyen et al., 2014) or diversity (Munson and Resnick, 2010) and hint at the existence of a variety of attitudes (Karakayali et al., 2018). They command further research on the expectations and literacy of users toward algorithmic devices. Furthermore, robots or "bots" (Ferrara et al., 2016), constitute a hybrid class of actors at the intersection of algorithms and users. They imitate human agency and thus contribute to socio-semantic dynamics as non-human actors *on par* with humans. While they may sometimes positively contribute to social cognition processes (for example algorithmic governance in Wikipedia, see Niederer and van Dijck, 2010; Müller-Birn et al., 2013), they are often studied for their possible role in distorting the socio-semantic dynamics of digital spaces (Shao et al., 2018), with sometimes concrete real-world effects e.g., tampering with stock markets or elections (Thomas et al., 2012).

Digital studies. This research more broadly connects with the history and epistemology of so-called digital studies. On the social science side, there has been a slow yet steady recognition of online communities as a legitimate investigation field *per se* — "a sense of the Internet as simply another context where social life is lived, where research methods are applied, and where contemporary social issues are addressed", as Hine (2004) nicely put it. This may be originally attributed to two non-exclusive movements: first, the progressive use of electronic devices to "digitize" the classical toolbox of conventional field research (Murthy, 2008; Ruppert et al., 2013) and, second and most importantly, the construal of computer networks as something essentially social (Flichy, 2000; Wellman, 2001), following the vision of Licklider and Taylor (1968) where computers should, more than anything, be human communication devices facilitating distributed cognition within groups of common interest. Social science scholarship has increasingly focused on the integration of online communication in everyday life, and at the everyday life in online communities, including seminal and general studies on virtual societies (Rheingold, 1993; Turner, 2006) and the specificities of this research (Hine, 2000; Wilson and Peterson, 2002). As a sizable share of my work has dealt with ICT platforms, at the interface between formal and social sciences, I explored in Roth (2019b) the epistemological settings and actors linked to three different acceptations of the term "digital humanities" (DH), distinguishing "humanities of the digital" (on online communities) from "digitized humanities" (on digital corpuses) and "numeric humanities" (dealing with mathematical models). I could show that the "DH" and "CSS" labels correspond to two markedly distinct epistemic communities: one linked to a certain tradition in human sciences paying a special attention to corpuses and their conservation, the other focused on more sociological issues. Online socio-technical systems offer a perfect playground at the interface between "humanities of the digital" and "numerical humanities", further explaining the tremendous interest of CSS in uniting both research objects.

Future perspectives

My research activity so far has essentially focused on (i) socio-semantic modeling frameworks, including hypergraphs and lattices, (ii) socio-technical systems, including science and online communities, and (iii) the underlying structure and dynamics of graphs in all generality. These questions relate to the broad issue of social morphogenesis: "how to develop an adequate theoretical account which deals simultaneously with men constituting society and the social formation of human agents" (Archer, 1982). Previous contributions may also now be modestly framed as a preliminary for the empirical understanding of the socio-semantic systems at a meta-level — by meta-level I am referring to the meta-structure i.e., the various shapes these systems may take. In effect, the present manuscript shall have shown that the present state of the art already sheds light on many of the key pieces of this ambitious puzzle: morphogenesis processes at the micro- or meso-level, epistemic community structure at the meso- or macro-level, co-evolution, diffusion, influence, authority at the micro- and macro-level — first and foremost in the context of online media and, to a distinct extent, scientific communities; two socio-semantic systems which variously contribute to shape our everyday epistemic landscape. Many of these phenomena have all been understood in a multitude of contexts and a variety of ways which sometimes yield partly divergent results. In other words, a truly integrated picture might still be missing: understanding the diversity of socio-semantic diversity might require the unification of several of the above-mentioned research streams.

As we have seen, socio-semantic hypergraphs provide a meta-framework that accommodates a vast majority of the existing hybrid formalisms of the social and semantic structure. Just the same, we should endeavor at developing meta-structural tools to accommodate for the variety of results showing that, under such and such circumstances (depending on topics, on link types, on time), there is socio-semantic cohesiveness, whereas there is none, or less, under other circumstances — some apparently conflicting observations from the state of the art possibly call for new hyperparameters that explain this meta-diversity. To this end, some of the most immediate research issues include the description of socio-semantic diversity dynamics at the meso-macro level of communities, by studying the presence, absence, emergence, stability, disappearance and more broadly the relative dynamic configuration of socio-semantic clusters. Most existing works focus on social phylogenies (i.e., dynamics and evolution of the social clusters), possibly projecting concepts on social hyperedges, or on semantic phylogenies, possibly projecting actors on semantic hyperedges, or they focus on a single case study that might blur the identification of regularities and diversity at a meta-level. Put differently, we might need a typology of the various socio-semantic configurations in hypergraphic and dynamic terms, of the transitions between such and such configuration, and, perhaps most importantly, of the drivers of these transitions and of the evolution of these drivers across time and space. Further, this would provide the background against which to study actor-centered dynamics and patterns of serendipity or conformism. For instance, the identification of pockets of nodes whose opinions are both strongly similar (relative increase of coherence, or decrease of diversity) and markedly distinct from the "local" average (precisely in the sense of a mean background landscape) could lead to a socio-cognitive theory of socio-semantic systems able to distinguish core/periphery configurations from polycentric ones and, again, transitions between them.

The meta-diversity should, again, be understood both in structural and semantic terms. The former focuses on the social network: can we identify typical, recurrent network shapes, are there particular interaction modes (e.g., distinct homophilic behaviors in some parts or types of networks)? The latter deals with the discursive content and partly rely on advances in natural language processing, beyond distributional approaches (Menezes and Roth, 2019b): are there specific terms or beliefs, rhetorical elements or narratives which are typical of certain subclusters? Is it possible to identify hyperparameters for successful (or "fit") discourses or sets of interrelated discourses, which are further disconnected from the rest of the system? Are there specific attitudes or discourses towards more consensual claims (at least in terms of "mainstream" discourses with respect to a given system)? What does such or such configuration entail in terms of further interactions and discourses, receptivity to influences and topics external to some cluster, reinforcement mechanisms jointly affecting influence and selection? This enables a finer analysis of the micro-level: what are the various ways in which the immediate neighborhood of ego may change across time, is it possible to formally define categories of joint breakpoints in discourses and interactions for categories of individuals? Meta-diversity has a temporal aspect too: can we observe small numbers of typical, even natural, timescales in socio-semantic systems? This could for instance be characterized by different forms of blindness to the remote or recent past: what disappears quickly, do actors focus on a number of elements (neighbors, sources or topics) whose diversity and durability grows or decreases in a typical manner?

Of particular interest, finally, is the contribution of algorithms to the ordering and filtering of individuals and ideas in socio-semantic systems. One should not only think of online media, where ranking and recommendation algorithms have become ubiquitous: science, too, is filled with algorithms whose principles are not evidently transparent, for instance when accessing Google Scholar's computation results of h-indices for people or ranked query results for content. While there is a relatively long-running debate on the impact of metrics in academia (Weingart, 2005; Hirsch, 2007; Waltman and van Eck, 2012), the fate of scientific communities operating under such or such filtering/evaluation rule remains to be studied in an *integral* way (in the above sense), especially regarding the impact on socio-semantic diversity.

To summarize, an integrated and ambitious research program for the coming years should thus aim equally at studying the joint effect of human algorithms (i.e. "behaviors") and man-designed algorithms on the diversity of our socio-cognitive landscapes, integrating the various strands of research which could crucially, yet separately, as of today, contribute to this multi-dimensional and interdisciplinary endeavor. These efforts aim more broadly at establishing the empirical bases for upcoming cultural contagion models which would be able to both take into account micro-level cognitive phenomena and interaction and diffusion dynamics at higher levels.

References

Abbott, A., 1995. Things of Boundaries. *Social Research*, 62(4):857–882.

Abrahamson, E., Rosenkopf, L., 1997. Social Network Effects on the Extent of Innovation Diffusion: A Computer Simulation. *Organization Science*, 8(3):289–309.

Acar, E., Dunlavy, D.M., Kolda, T.G., 2009. Link prediction on evolving data using matrix and tensor factorizations. In *Proc. ICDMW'09 Intl. Conf. Data Mining Workshops*, 262–269. IEEE.

Adamic, L.A., Glance, N., 2005. The political blogosphere and the 2004 U.S. election: divided they blog. In *LinkKDD '05: Proc. 3rd Intl. Workshop on Link discovery*, 36–43. ACM.

Adar, E., Zhang, L., Adamic, L.A., Lukose, R.M., 2004. Implicit Structure and the Dynamics of Blogspace. In *Workshop on the Weblogging Ecosystem, 13th Intl. World Wide Web Conf.*

Agarwal, N., Liu, H., Tang, L., Yu, P.S., 2008. Identifying the Influential Bloggers in a Community. In *Proc. WSDM'08 Intl. Conf. on Web Search and Data Mining*, 207–218. ACM.

Ahn, Y.Y., Bagrow, J.P., Lehmann, S., 2010. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764.

Aiello, L.M., Barbieri, N., 2017. Evolution of Ego-networks in Social Media with Link Recommendations. In *Proc. WSDM'17 10th Intl. Conf. on Web Search and Data Mining*, 111–120. ACM.

Aiello, L.M., Barrat, A., Cattuto, C., Ruffo, G., Schifanella, R., 2010. Link creation and profile alignment in the aNobii social network. In *Proc. SocialCom'10 2nd Intl Conf on Social Computing*, 249–256. IEEE.

Aiello, W., Chung, F., Lu, L., 2000. A random graph model for massive graphs. In *Proc. 32nd Annual Symp. on Theory of computing*, 171–180. ACM.

Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P., 2008. Mixed membership stochastic blockmodels. *Journal of machine learning research*, 9:1981–2014.

Alba, R.D., 1973. A Graph-Theoretic Definition of a Sociometric Clique. *Journal of Mathematical Sociology*, 3:113–126.

Albert, R., Barabási, A.L., 2002. Statistical Mechanics of Complex Networks. *Rev Mod Phys*, 74:47–97.

Alexander, J., Smith, P., 2001. The Strong Program in Cultural Sociology: Elements of a Structural Hermeneutics. In J. Turner, ed., *Handbook of sociological theory*, 135–150. Springer, New York.

AlShebli, B.K., Rahwan, T., Woon, W.L., 2018. The preeminence of ethnic diversity in scientific collaboration. *Nature communications*, 9(1):1–10.

Amblard, F., Deffuant, G., 2004. The role of network topology on extremism propagation with the relative agreement opinion dynamics. *Physica A*, 343:725–738.

Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., Lalmas, M., 2020. Algorithmic effects on the diversity of consumption on spotify. In *Proc. WWW'20 The Web Conference*, 2155–2165. ACM.

Anderson, C.J., Wasserman, S., Crouch, B., 1999. A p^* primer: logit models for social networks. *Social Networks*, 21:37–66.

Anderson, C.J., Wasserman, S., Faust, K., 1992. Building stochastic blockmodels. *Social Networks*, 14:137–161.

Angeli, G., Premkumar, M.J.J., Manning, C.D., 2015. Leveraging linguistic structure for open domain information extraction. In *Proc. 53rd Annual Meeting Association Comp. Ling. and 7th Intl Joint Conf on NLP*, vol. 1: Long papers, 344–354. ACL.

Angelov, D., 2020. Top2Vec: Distributed Representations of Topics. *arXiv preprint arXiv:2008.09470.*

Aral, S., Muchnik, L., Sundararajan, A., 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *PNAS*, 106(51):21544–21549.

Archer, M.S., 1982. Morphogenesis versus structuration: on combining structure and action. *The British Journal of Sociology*, 33(4):455–483.

Arenas, A., Danon, L., Diaz-Guilera, A., Gleiser, P.M., Guimera, R., 2004. Community analysis in social networks. *Eur. Phys. Journal B*, 38(2):373–380.

Arora, V., Ventresca, M., 2017. Action-based Modeling of Complex Networks. *Scientific Reports*, 7(6673).

Asur, S., Huberman, B.A., 2010. Predicting the future with social media. In *2010 IEEE/WIC/ACM Intl. Conf. Web Intelligence and Intelligent Agent Technology*, vol. 1, 492–499. IEEE.

Axelrod, R., 1997. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226.

Axtell, R., Epstein, J.M., Young, H.P., 2001. The Emergence of Classes in a Multi-Agent Bargaining Model. In H.P. Young, S. Durlauf, eds., *Social Dynamics*, 191–211. Cambridge: MIT Press.

Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X., 2006. Group formation in large social networks: membership, growth, and evolution. In *Proc. KDD'06: 12th SIGKDD Intl. Conf. Knowledge Discovery and Data mining*, 44–54. ACM.

Bailey, A., Ventresca, M., Ombuki-Berman, B., 2012. Automatic generation of graph models for complex networks by genetic programming. In *Proc. GECCO'12 14th Annual Conference on Genetic and Evolutionary Computation*, 711–718. ACM.

Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J., 2011. Everyone's an influencer: quantifying influence on twitter. In *Proc. WSDM'11 4th Intl Conf Web Search & Data Mining*, 65–74. ACM.

Bakshy, E., Karrer, B., Adamic, L.A., 2009. Social influence and the diffusion of user-created content. In *Proc. EC'09 10th Conf on Electronic Commerce*, 325–334. ACM.

Bakshy, E., Messing, S., Adamic, L.A., 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132.

Baltzer, A., Karsai, M., Roth, C., 2019. Interactional and Informational Attention on Twitter. *Information*, 10(8):250.

Barabási, A.L., Albert, R., 1999. Emergence of Scaling in Random Networks. *Science*, 286:509–512.

Barabási, A.L., Jeong, H., Ravasz, E., Neda, Z., Vicsek, T., Schubert, A., 2002. Evolution of the social network of scientific collaborations. *Physica A*, 311:590–614.

Barber, M.J., 2007. Modularity and community detection in bipartite networks. *Physical Review E*, 76(066102).

Barberá, P., 2020. Social media, echo chambers, and political polarization. *Social Media and Democracy: The State of the Field, Prospects for Reform*, 34.

Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., Bonneau, R., 2015. Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, 26(10):1531–1542.

Barbut, M., Monjardet, B., 1970. *Algèbre et Combinatoire*, vol. II. Hachette, Paris.

Barthélemy, M., 2011. Spatial Networks. *Physics Reports*, 499:1–101.

Basov, N., Breiger, R., Hellsten, I., 2020. Socio-semantic and other dualities. *Poetics*, 78:101433.

Basov, N., Brennecke, J., 2017. Duality beyond dyads: Multiplex patterning of social ties and cultural meanings. In *Structure, Content and Meaning of Organizational Networks: Extending Network Thinking, Research in the Sociology of Organizations*, vol. 53, 87–112. Emerald Publishing.

Basov, N., de Nooy, W., Nenko, A., 2019. Local meaning structures: mixed-method sociosemantic network analysis. *American Journal of Cultural Sociology*.

Bastard, I., Cardon, D., Charbey, R., Cointet, J.P., Prieur, C., 2017. What do we do on Facebook? Activity patterns and relational structures on a social network. *Sociologie*, 1(8):57–82.

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.G., Petri, G., 2020. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, 874:1–92.

Beiró, M.G., Panisson, A., Tizzoni, M., Cattuto, C., 2016. Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Science*, 5:1–15.

Ben Jdidia, M., Robardet, C., Fleury, E., 2007. Communities detection and analysis of their dynamics in collaborative networks. In *Proc 2nd ICDIM Intl Conf Digital Information Management*, 744–749.

Berger, N., Borgs, C., Chayes, J., D'Souza, R., Kleinberg, R., 2004. Competition-Induced Preferential Attachment. In *Proc. 31st ICALP'04 Intl. Colloquium on Automata, Languages and Programming*, *LNCS*, vol. 3142, 208–221. Springer.

Blau, P.M., 1970. A formal theory of differentiation in organizations. *American sociological review*, 35(2):201–218.

Blei, D.M., 2012. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

—, 2014. Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1:203–232.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.

Bliss, C.A., Frank, M.R., Danforth, C.M., Dodds, P.S., 2014. An evolutionary algorithm approach to link prediction in dynamic social networks. *Journal of Computational Science*, 5(5):750–764.

Block, P., Stadtfeld, C., Snijders, T.A.B., 2019. Forms of Dependence: Comparing SAOMs and ERGMs From Basic Principles. *Sociological Methods & Research*, 48(1):202–239.

Boeck, P.D., Rosenberg, S., 1988. Hierarchical Classes: Model and Data Analysis. *Psychometrika*, 53(3):361–381.

Boguna, M., Pastor-Satorras, R., 2003. Class of correlated random networks with hidden variables. *Physical Review E*, 68:036112.

Boguna, M., Pastor-Satorras, R., Diaz-Guilera, A., Arenas, A., 2004. Models of social networks based on social distance attachment. *Physical Review E*, 70:056122.

Borge-Holthoefer, J., Rivero, A., Moreno, Y., 2012. Locating privileged spreaders on an online social network. *Physical review E*, 85(6):066123.

Börner, K., Chen, C., Boyack, K.W., 2003. Visualizing Knowledge Domains. *Annual review of information science and technology*, 37(1):179–255.

Bourdieu, P., 1985. The Social Space and the Genesis of Groups. *Theory & Society*, 14(6):723–744.

Bouveyron, C., Latouche, P., Zreik, R., 2018. The stochastic topic block model for the clustering of vertices in networks with textual edges. *Statistics and Computing*, 28(1):11–31.

Brandes, U., Lerner, J., 2008. Visual analysis of controversy in user-generated encyclopedias. *Information visualization*, 7(1):34–48.

Brandes, U., Raab, J., Wagner, D., 2001. Exploratory network visualization: Simultaneous display of actor status and connections. *Journal of Social Structure*, 2(4):1–28.

Breiger, R.L., 1974. The Duality of Persons and Groups. *Social Forces*, 53(2):181–190.

—, 2010. Dualities of culture and structure: Seeing through cultural holes. In J. Fuhse, S. Mützel, eds., *Relationale Soziologie*, 37–47. Springer.

Breiger, R.L., Boorman, S.A., Arabie, P., 1975. An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12(3):328–383.

Brennecke, J., Rank, O.N., 2016. The interplay between formal project memberships and informal advice seeking in knowledge-intensive firms: A multilevel network approach. *Social Networks*, 44:307–318.

Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. *Nature*, 439:462–465.

Bruns, A., Sauter, T., 2015. Anatomie eines Trending Topics: Methoden zur Visualisierung von Retweet-Ketten. In A. Maireder, J. Ausserhofer, C. Schumann, M. Taddicken, eds., *Digitale Methoden in der Kommunikationswissenschaft*, 141–161.

Bryant, S.L., Forte, A., Bruckman, A., 2005. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In *Proc. 2005 Intl. SIGGROUP Conf. Supporting group work*, 1–10. ACM.

Burt, R.S., 1987. Social Contagion and Innovation: Cohesion Versus Structural Equivalence. *American Journal of Sociology*, 92(6):1287–1335.

—, 2004. Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399.

Caldarelli, G., Capocci, A., Rios, P.D.L., Munoz, M.A., 2002. Scale-Free Networks from Varying Vertex Intrinsic Fitness. *Phys. Rev. Lett.*, 89(25):258702.

Callon, M., 2006. Can methods for analysing large numbers organize a productive dialogue with the actors they study? *European Management Review*, 3(1):7–16.

Callon, M., Courtial, J.P., Laville, F., 1991. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22(1):155–205.

Callon, M., Law, J., Rip, A., 1986. *Mapping the dynamics of science and technology*. MacMillan Press.

Cambrosio, A., Keating, P., Mogoutov, A., 2004. Mapping collaborative work and innovation in biomedicine. *Social Studies of Science*, 34(3):325–364.

Cardon, D., Fouetillou, G., Roth, C., 2011. Two paths of glory – Structural positions and trajectories of websites within their topical territory. In *Proc. ICWSM 5th Intl Conf Weblogs Social Media*, 58–65. AAAI.

—, 2014. Topographie de la renommée en ligne : un modèle structurel des communautés thématiques du web français et allemand. *Réseaux*, 188:85–119.

Carley, K.M., 1986. An Approach for Relating Social Structure to Cognitive Structure. *Journal of Mathematical Sociology*, 12(2):137–189.

—, 1994. Extracting culture through textual analysis. *Poetics*, 22(4):291–312.

—, 2002. Smart agents and organizations of the future. In L. Lievrouw, S. Livingstone, eds., *Handbook of new media*, vol. 12, 206–220. Sage.

Castellano, C., Fortunato, S., Loreto, V., 2009. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646.

Cattuto, C., Loreto, V., Pietronero, L., 2007. Semiotic Dynamics and Collaborative Tagging. *PNAS*, 104(1461–1464).

Centola, D., Gonzalez-Avella, J.C., Eguiluz, V.M., San Miguel, M., 2007. Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, 51(6):905–929.

Centola, D., Macy, M., 2007. Complex Contagions and the Weakness of Long Ties. *American Journal of Sociology*, 113:702–734.

Cerinšek, M., Batagelj, V., 2015. Generalized two-mode cores. *Social Networks*, 42:80–87.

Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P., 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proc. ICWSM'10 4th Intl. Conf. Weblogs and Social Media*, 10–17. AAAI.

Cha, M., Mislove, A., Gummadi, K.P., 2009a. A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. In *Proc. 18th WWW Intl Conf World Wide Web*, 721–730. ACM.

Cha, M., Perez, J.A.N., Haddadi, H., 2009b. Flash Floods and Ripples: The Spread of Media Content through the Blogosphere. In *Proc. ICWSM 3rd Intl. Conf. Weblogs and Social Media*, 8–15. AAAI.

Chang, J., Blei, D., 2009. Relational topic models for document networks. *PLMR*, 5:81–88.

Chavalarias, D., Cointet, J.P., 2013. Phylomemetic Patterns in Science Evolution – The Rise and Fall of Scientific Fields. *PLoS ONE*, 8(2):e54847.

Chavalarias, D., J.-P., C., Cornilleau, L., K., D.T., Mogoutov, A., Roth, C., Savy, T., 2011. Stream of media issues: Approaches for monitoring world food security. http://pulseweb.cortext.net/static/files/wp1.pdf.

Chavalarias, D., Wallach, J.D., Li, A.H.T., Ioannidis, J.P., 2016. Evolution of reporting *P* values in the biomedical literature, 1990-2015. *JAMA*, 315(11):1141–1148.

Chen, C.C., Roth, C., 2010. Motivations and Constraints to Member Engagement in Flickr Groups. In *Quality in Socio-Techno Systems (QTESO) Workshop, IEEE Conf. on Self-Adaptive and Self-Organizing Systems (SASO)*, 126–130.

—, 2011. The role of (non-)conformism in rating platforms. In *Proc SocialCom'11 3rd Intl Conf on Social Computing*, 347–353. IEEE.

—, 2012. {{Citation needed}}: The dynamics of referencing in Wikipedia. In *Proc. WikiSym'12 8th Annual Intl Symposium on Wikis and Open Collaboration*, 8, 1–4. ACM.

Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J., 2014. Can cascades be predicted? In *Proc. WWW'14 23rd Intl. Conf. on World Wide Web*, 925–936. ACM.

Chikhaoui, B., Chiazzaro, M., Wang, S., Sotir, M., 2017. Detecting communities of authority and analyzing their influence in dynamic social networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(6):1–28.

Cho, E., Myers, S.A., Leskovec, J., 2011. Friendship and Mobility: User Movement in Location-based Social Networks. In *Proc. KDD'11: 17th SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, 1082–1090. ACM.

Chodrow, P.S., 2020. Configuration models of random hypergraphs. *Journal of Complex Networks*, 8(3):cnaa018.

Christakis, N.A., Fowler, J.H., 2013. Social contagion theory: examining dynamic social networks and human behavior. *Statistics in medicine*, 32(4):556–577.

Ciampaglia, G.L., Ferrara, E., Flammini, A., 2014. Collective behaviors and networks. *EPJ Data Science*, 3(37).

Cinelli, M., Morales, G.D.F., Galeazzi, A., Quattrociocchi, W., Starnini, M., 2021. The echo chamber effect on social media. *PNAS*, 118(9):e2023301118.

Claidière, N., Sperber, D., 2007. Commentary: The role of attraction in cultural evolution. *Journal of Cognition and Culture*, 7:89–111.

Clauset, A., Moore, C., Newman, M.E.J., 2008. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453:98–101.

Cointet, J.P., Faure, E., Roth, C., 2007. Intertemporal topic correlations in online media. In *Proc. ICWSM Intl Conf Weblogs Social Media*, 38.

Cointet, J.P., Roth, C., 2007a. How Realistic Should Knowledge Diffusion Models Be? *Journal of Artificial Societies and Social Simulation*, 10(3):5.

—, 2007b. Information diffusion in realistic networks. In *Actes d'AlgoTel 9e rencontres francophones sur les aspects ALGOrithmiques des TELécommunications, Ile d'Oléron*, 4 p. Best Student Paper.

—, 2009. Socio-semantic dynamics in a blog network. In A. Pentland, J. Zhan, eds., *Proc. Intl. Conf. Computational Science and Engineering*, vol. 4, 114–121. IEEE. Best Paper Award.

—, 2010. Local Networks, Local Topics: Structural and Semantic Proximity in Blogspace. In *Proc. 4th ICWSM Intl. Conf. on Weblogs and Social Media*, 223–226. AAAI.

Colizza, V., Banavar, J.R., Maritan, A., Rinaldo, A., 2004. Network Structures from Selection Principles. *Physical Review Letters*, 92(19):198701.

Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., Menczer, F., 2011. Political Polarization on Twitter. In *Proc. 5th ICWSM Intl. Conf. Weblogs and Social Media*, 89–96. AAAI.

Conte, R., 2000. Memes through (social) minds. In R. Aunger, ed., *Darwinizing Culture: The Status of Memetics as a Science*, 83–120. Oxford: Oxford University Press.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.P., Sanchez, A., Nowak, A., Flache, A., Miguel, M.S., Helbing, D., 2012. Manifesto of computational social science. *European Physical Journal Special Topics*, 214(1):325–346.

Corominas-Murtra, B., Goñi, J., Solé, R.V., Rodríguez-Caso, C., 2013. On the origins of hierarchy in complex networks. *PNAS*, 110(33):13316–13321.

Cowan, R., Jonard, N., 2004. Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, 28:1557–1575.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S., 2008. Feedback Effects Between Similarity and Social Influence in Online Communities. In *Proc. KDD'08 SIGKDD 14th Intl Conf on Knowledge Discovery and Data mining*, 160–168. ACM.

Crane, R., Sornette, D., 2008. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS*, 105(41):15649–15653.

Crépey, P., Alvarez, F.P., Barthélemy, M., 2006. Epidemic variability in complex networks. *Physical Review E*, 73(4):046131.

Cruz, J.D., Bothorel, C., Poulet, F., 2013. Community Detection and Visualization in Social Networks: Integrating Structural and Semantic Information. *ACM Transactions on Intelligent Systems and Technology*, 5(1):11.

Cucchiarelli, A., D'Antonio, F., Velardi, P., 2010. Analyzing collaborations through content-based social networks. In *Computational Social Network Analysis*, 387–409. Springer.

Cui, W., Liu, S., Tan, L., Shi, C., Song, Y., Gao, Z., Qu, H., Tong, X., 2011. TextFlow: Towards Better Understanding of Evolving Topics in Text. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2412–2421.

Cummings, J., Kiesler, S., 2007. Coordination Costs and Project Outcomes in Multi-University Collaborations. *Research Policy*, 36(10):1620–1634.

da Fontoura Costa, L., Rodrigues, F.A., Travieso, G., Boas, P.R.V., 2007. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242.

Dahlberg, L., 2001. Computer-Mediated Communication and The Public Sphere: A Critical Analysis. *Journal of Computer-Mediated Communication*, 7(1):JCMC714.

Datta, A., Yong, J.T.T., Braghin, S., 2014. The zen of multidisciplinary team recommendation. *Journal of the Association for Information Science and Technology*, 65(12):2518–2533.

Davey, B.A., Priestley, H.A., 2002. *Introduction to Lattices and Order*. Cambridge, UK: Cambridge University Press, 2nd ed.

Davison, B.D., 2000. Topical locality in the web. In *Proc. 23rd SIGIR Annual Intl Conf on Research and development in information retrieval*, 272–279. ACM.

Dawkins, R., 1976. *The Selfish Gene*, chap. 11: Memes, The New Replicator, 189–201. Oxford: Oxford University Press.

de Solla Price, D.J., 1976. A General Theory of Bibliometric and Other Cumulative Advantage Processes. *Journal of the American Society for Information Science*, 27(5–6):292–306.

deB. Beaver, D., 1986. Collaboration and Teamwork in Physics. *Czech Journal of Physics B*, 36(14–18).

Deffuant, G., Neau, D., Amblard, F., Weisbuch, G., 2000. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98.

Deroian, F., 2002. Formation of social networks and diffusion of innovations. *Research Policy*, 31:835–846.

Diesner, J., Carley, K., 2005. Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a Novel Method for Network Text Analysis. In *Causal mapping for research in information technology*, 81–108. IGI Global.

Ding, Y., 2011. Community detection: Topological vs. topical. *Journal of Informetrics*, 5(4):498–514.

Doreian, P., 1979. On the Evolution of Group and Network Structure. *Social Networks*, 2:235–252.

Doreian, P., Batagelj, V., Ferligoj, A., 2004. Generalized blockmodeling of two-mode network data. *Social Networks*, 26(1):29–53.

Dörk, M., Riche, N.H., Ramos, G., Dumais, S., 2012. PivotPaths: Strolling through Faceted Information Spaces. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2709–2718.

Dorogovtsev, S.N., Mendes, J.F.F., 2000. Evolution of networks with aging of sites. *Physical Review E*, 62:1842–1845.

Dourish, P., Chalmers, M., 1994. Running out of space: Models of information navigation. In *Proc HCI Human-Computer Interaction*, vol. 94, 23–26.

D'Souza, R.M., Borgs, C., Chayes, J.T., Berger, N., Kleinberg, R.D., 2007. Emergence of tempered preferential attachment from optimization. *PNAS*, 104:6112–6117.

Eguiluz, V.M., Klemm, K., 2002. Epidemic threshold in structured scale-free networks. *Physical Review Letters*, 89:108701.

Eleta, I., Golbeck, J., 2014. Multilingual use of Twitter: Social networks at the language frontier. *Computers in Human Behavior*, 41:424–432.

Elgesem, D., Stskal, L., Diakopoulos, N., 2015. Structure and content of discourse on climate change in the blogosphere: The big picture. *Environmental Communication*, 9(2):169–188.

Ellison, G., Fudenberg, D., 1995. Word-of-Mouth Communication and Social Learning. *Quarterly Journal of Economics*, 110(1):93–125.

Emirbayer, M., Goodwin, J., 1994. Network analysis, culture, and the problem of agency. *American Journal of Sociology*, 99(6):1411–1454.

Epstein, R., Robertson, R.E., 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *PNAS*, 112(33):E4512–E4521.

Erdős, P., Rényi, A., 1959. On random graphs. *Publicationes Mathematicae*, 6:290–297.

Estrada, E., 2007. Topological structural classes of complex networks. *Physical Review E*, 75(1):016103.

Etling, B., Kelly, J., Faris, R., Palfrey, J., 2010. Mapping the Arabic blogosphere: Politics and dissent online. *New Media & Society*, 12(8):1225–1243.

Evans, J.A., Aceves, P., 2016. Machine translation: Mining text for social theory. *Annual Review of Sociology*, 42:21–50.

Everett, M.G., Borgatti, S.P., 2013. The dual-projection approach for two-mode networks. *Social Networks*, 35(2):204–210.

Fabrikant, A., Koutsoupias, E., Papadimitriou, C.H., 2002. Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet. In *Proc. 29th ICALP Intl. Coll. on Automata, Languages and Programming*, 110–122. Springer.

Ferguson, J.E., Groenewegen, P., Moser, C., Borgatti, S.P., Mohr, J.W., 2017. Structure, Content and Meaning of Organizational Networks: Extending Network Thinking. *Research in the Sociology of Organizations*, 53:1–15.

Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A., 2016. The Rise of Social Bots. *Communications of the ACM*, 59(7):96–104.

Fienberg, S.E., Meyer, M.M., Wasserman, S.S., 1985. Statistical Analysis of Multiple Sociometric Relations. *J. Am. Stat. Association*, 80(389):51–67.

Fine, G.A., Kleinman, S., 1983. Network and Meaning: An Interactionist Approach to Structure. *Symbolic Interaction*, 6(1):97–110.

Finin, T., Ding, L., Zhou, L., Joshi, A., 2005. Social networking on the semantic web. *The Learning Organization*, 12(5):418–435.

Flache, A., Macy, M.W., 2011. Small worlds and cultural polarization. *Journal of Mathematical Sociology*, 35(1-3):146–176.

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., Lorenz, J., 2017. Models of Social Influence: Towards the Next Frontiers. *JASSS*, 20(4):2.

Flake, G.W., Lawrence, S., Giles, C.L., Coetzee, F.M., 2002. Self-organization and identification of web communities. *Computer*, 35(3):66–70.

Flichy, P., 2000. Internet or the Ideal Scientific Community. *Réseaux*, 7(2):155–182.

Forte, A., Andalibi, N., Gorichanaz, T., Kim, M.C., Park, T., Halfaker, A., 2018. Information fortification: An online citation behavior. In *Proc Group'18 Intl Conf Supporting Groupwork*, 83–92. ACM.

Fortunato, S., 2010. Community detection in graphs. *Physics Reports*, 486:75—174.

Fortunato, S., Barthelemy, M., 2007. Resolution limit in community detection. *PNAS*, 104(1):36–41.

Fortunato, S., Flammini, A., Menczer, F., 2006. Scale-free network growth by ranking. *Physical Review Letters*, 96:218701.

Frank, O., Strauss, D., 1986. Markov Graphs. *J. Am. Stat. Association*, 81(395):832–842.

Freeman, L.C., White, D.R., 1993. Using Galois Lattices to Represent Network Data. *Sociological Methodology*, 23:127–146.

Friggeri, A., Adamic, L., Eckles, D., Cheng, J., 2014. Rumor Cascades. In *Proc. 8th ICWSM Intl. Conf. Weblogs Social Media*, 101–110. AAAI.

Fuhse, J., 2009. The Meaning Structure of Social Networks. *Sociological Theory*, 27(1):51–73.

Fuhse, J., Stuhler, O., Riebling, J., Martin, J.L., 2020. Relating social and symbolic relations in quantitative text analysis. A study of parliamentary discourse in the Weimar Republic. *Poetics*, 78:101363.

Ganesh, A., Massoulié, L., Towsley, D., 2005. The effect of network topology on the spread of epidemics. In *Proc. 24th Annual Joint Conf IEEE Computer and Communications Societies*, vol. 2, 1455–1466. IEEE.

Ganter, B., Wille, R., 1999. *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin.

Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M., 2016. Quantifying Controversy in Social Media. In *Proc. 9th WSDM Intl Conf Web Search and Data Mining*, 33–42. ACM.

Garimella, K., Gionis, A., Morales, G.D.F., Mathioudakis, M., 2018. Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. In *Proc. WWW'18 Intl Conf World Wide Web*, 913–922. ACM.

Gaumont, N., Panahi, M., Chavalarias, D., 2018. Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election. *PLOS ONE*, 13(9):e0201879.

Gerlach, M., Peixoto, T.P., Altmann, E.G., 2018. A network approach to topic models. *Science advances*, 4(7):eaaq1360.

Gibson, D., Kleinberg, J., Raghavan, P., 2000. Clustering categorical data: An approach based on dynamical systems. *VLDB Journal*, 8(3):222–236.

Gilbert, N., 1997. A Simulation of the Structure of Academic Science. *Soc. Res. Online*, 2(2):1–15.

Gilbert, N., Troitzsch, K.G., 1999. *Simulation for the Social Scientist*. Open University Press.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. *Nature*, 457:1012–1014.

Gkantsidis, C., Mihail, M., Zegura, E.W., 2003. The Markov Chain Simulation Method for Generating Connected Power Law Random Graphs. In *Proc. 5th Workshop on Algorithm Engineering and Experiments (ALENEX)*, 16–25.

Gloor, P.A., Krauss, J., Nann, S., Fischbach, K., Schoder, D., 2009. Web Science 2.0: Identifying Trends through Semantic Social Network Analysis. In *Proc. Intl. Conf. Computational Science and Engineering*, vol. 4, 215–222. IEEE.

Godart, F.C., Galunic, C., 2019. Explaining the Popularity of Cultural Elements: Networks, Culture, and the Structural Embeddedness of High Fashion Trends. *Organization Science*, 30(1):151–168.

Godart, F.C., White, H.C., 2010. Switchings under uncertainty: The coming and becoming of meanings. *Poetics*, 38:567–586.

Goel, S., Anderson, A., Hofman, J., Watts, D.J., 2016. The structural Virality of Online Diffusion. *Management Science*, 62(1):180–196.

Goetz, M., Leskovec, J., McGlohon, M., Faloutsos, C., 2009. Modeling Blog Dynamics. In *Proc. 3rd ICWSM Intl Conf Weblogs Social Media*, 26–33. AAAI.

Goldberg, A., Srivastava, S.B., Manian, V.G., Monroe, W., Potts, C., 2016. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *American Sociological Review*, 81(6):1190–1222.

Goñi, J., Avena-Koenigsberger, A., de Mendizabal, N.V., van den Heuvel, M.P., Betzel, R.F., Sporns, O., 2013. Exploring the Morphospace of Communication Efficiency in Complex Networks. *PLOS ONE*, 8(3):e58070.

González, M.C., Hidalgo, C.A., Barabási, A.L., 2008. Understanding individual human mobility patterns. *Nature*, 453:779–782.

González-Avella, J.C., Cosenza, M.G., Klemm, K., Eguíluz, V., Miguel, M.S., 2007. Information Feedback and Mass Media Effects in Cultural Dynamics. *JASSS*, 10(3):9.

Gonzalez-Bailon, S., Borge-Holthoefer, J., Rivero, A., Moreno, Y., 2011. The Dynamics of Protest Recruitment through an Online Network. *Scientific Reports*, 1(197):1–7.

Goyal, P., Ferrara, E., 2018. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 151:78–94.

Goyal, S., 2005. Learning in networks: A survey. In G. Demange, M. Wooders, eds., *Group formation in economics: Networks, clubs, and coalitions*, chap. 4, 122–167. Cambridge University Press.

Grabowicz, P., Ramasco, J.J., Moro, E., Pujol, J.M., Eguiluz, V.M., 2012. Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media. *PLOS ONE*, 7(1):e29358.

Granovetter, M., 1978. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443.

Grim, P., 2009. Threshold Phenomena in Epistemic Networks. In *Papers from the AAAI Fall Symposium: Complex Adaptive Systems and the Threshold Effect*, 53–60.

Grimmer, J., Stewart, B.M., 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297.

Gross, T., Blasius, B., 2008. Adaptive coevolutionary networks: a review. *Journal of The Royal Society Interface*, 5:259–271.

Grover, A., Leskovec, J., 2016. node2vec: Scalable feature learning for networks. In *Proc. 22nd SIGKDD Intl Conf Knowledge discovery and data mining*, 855–864. ACM.

Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A., 2004. Information Diffusion Through Blogspace. In *Proc. WWW'04 13th Intl Conf on World Wide Web*, 491–501. ACM.

Guimerà, R., Sales-Pardo, M., 2009. Missing and spurious interactions and the reconstruction of complex networks. *PNAS*, 106(52):22073–22078.

Guimerà, R., Sales-Pardo, M., Amaral, L.A.N., 2007. Classes of complex networks defined by role-to-role connectivity profiles. *Nature Physics*, 3:63–69.

Guimerà, R., Sales-Pardo, M., Amaral, L.A.N., 2007. Module identification in bipartite and directed networks. *Physical Review E*, 76(036102).

Guimera, R., Uzzi, B., Spiro, J., Amaral, L.A.N., 2005. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science*, 308:697–702.

Haas, P., 1992. Introduction: epistemic communities and international policy coordination. *International Organization*, 46(1):1–35.

Hamberger, K., Houseman, M., White, D., 2011. Kinship network analysis. In P. Carrington, J. Scott, eds., *The Sage Handbook of Social Network Analysis*, 533–549. Sage.

Han, Y., Zhou, B., Pei, J., Jia, Y., 2009. Understanding importance of collaborations in co-authorship networks: A supportiveness analysis approach. In *Proc. SIAM Intl Conf Data Mining*, 1112–1123.

Hanneke, S., Fu, W., Xing, E.P., 2010. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.

Harrison, K.R., Ventresca, M., Ombuki-Berman, B.M., 2016. A meta-analysis of centrality measures for comparing and generating complex network models. *Journal of computational science*, 17:205–215.

Hasan, M.A., Chaoji, V., Salem, S., Zaki, M., 2006. Link prediction using supervised learning. In *Proc. SDM'06: Workshop on Link Analysis, Counterterrorism and Security*, vol. 30, 798–805.

Hasan, M.A., Zaki, M.J., 2011. A Survey of Link Prediction in Social Networks. In C. Aggarwal, ed., *Social Network Data Analytics*, 243–275.

Hegselmann, R., Krause, U., 2015. Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model. *Networks & Heterogeneous Media*, 10(3):477–509.

Hellsten, I., Leydesdorff, L., 2020. Automated analysis of actor-topic networks on twitter: New approaches to the analysis of socio-semantic networks. *JASIST*, 71(1):3–15.

Hethcote, H.W., 2000. The Mathematics of Infectious Diseases. *SIAM Review*, 42(4):599–653.

Himelboim, I., McCreery, S., Smith, M., 2013. Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication*, 18(2):154–174.

Himelboim, I., Smith, M.A., Rainie, L., Shneiderman, B., Espina, C., 2017. Classifying Twitter Topic-Networks Using Social Network Analysis. *Social Media + Society*, 3(1):1–13.

Hindman, M., Tsioutsiouliklis, K., Johnson, J.A., 2003. Googlearchy: How a Few Heavily-Linked Sites Dominate Politics on the Web. In *Proc. Annual Meeting of the Midwest Political Science Association*, 1–33.

Hine, C., 2000. *Virtual Ethnography*. Sage, London.

—, 2004. Social Research Methods and the Internet: A Thematic Review. *Sociological Research Online*, 9(2).

Hirsch, J.E., 2007. Does the *h* index have predictive power? *PNAS*, 104(49):19193–19198.

Hoffmann, I., 2014. The Populist Networks. spotlight Europe 2014/02, Bertelsmann Stiftung.

Holland, J.H., 1996. *Hidden order: How adaptation builds complexity*. Addison Wesley Longman.

Holland, P.W., Laskey, K.B., Leinhardt, S., 1983. Stochastic Blockmodels: First Steps. *Social Networks*, 5:109–137.

Holland, P.W., Leinhardt, S., 1977. A Dynamic Model for Social Networks. *Journal of Mathematical Sociology*, 5:5–20.

—, 1981. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373):33–65.

Holme, P., Kim, B.J., 2002. Growing Scale-Free Networks with Tunable Clustering. *Physical Review E*, 65:026107.

Holme, P., Newman, M.E.J., 2007. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E*, 74(056108).

Hopcroft, J., Khan, O., Kulis, B., Selman, B., 2004. Tracking evolving communities in large linked networks. *PNAS*, 101(suppl 1):5249–5253.

Hours, H., Fleury, E., Karsai, M., 2016. Link prediction in the twitter mention network: impacts of local structure and similarity of interest. In *Proc. 16th ICDMW Intl Conf Data Mining Workshops*, 454–461. IEEE.

Hutchins, E., 1995. *Cognition in the Wild*. MIT press.

Iribarren, J.L., Moro, E., 2011. Affinity paths and information diffusion in social networks. *Social networks*, 33(2):134–142.

Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G., 2008. Discovering shared conceptualizations in folksonomies. *Journal of Web Semantics*, 6(1):38–53.

Java, A., Song, X., Finin, T., Tseng, B., 2007. Why We Twitter: Understanding Microblogging Usage and Communities. In *Proc. 9th WebKDD / 1st SNA-KDD Workshop on Web Mining and Social Network Analysis*, 56–65. ACM.

Jones, B.F., Wuchty, S., Uzzi, B., 2008. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science*, 322(5905):1259–1262.

Kaiser, J., Puschmann, C., 2017. Alliance of antagonism: Counterpublics and polarization in online climate change communication. *Communication and the Public*, 2(4):371–387.

Karakayali, N., Kostem, B., Galip, I., 2018. Recommendation systems as technologies of the self: Algorithmic control and the formation of music taste. *Theory, Culture & Society*, 35(2):3–24.

Karpf, D., 2008. Understanding Blogspace. *Journal of Information Technology & Politics*, 5(4):369–385.

Karrer, B., Newman, M.E.J., 2010. Random graphs containing arbitrary distributions of subgraphs. *Physical Review E*, 82:066118.

Keegan, B.C., Gergle, D., Contractor, N., 2012. Staying in the loop: structure and dynamics of Wikipedia's breaking news collaborations. In *Proc. WikiSym'12 8th Annual Intl Symposium on Wikis and Open Collaboration*, 1, 1–10. ACM.

Kempe, D., Kleinberg, J., Tardos, E., 2003. Maximizing the spread of influence through a social network. In *Proc. KDD'03 9th SIGKDD Knowledge discovery and data mining*, 137–146. ACM.

Kim, S., Weber, I., Wei, L., Oh, A., 2014. Sociolinguistic Analysis of Twitter in Multilingual Societies. In *Proc. HT'14 25th Intl Conf Hypertext and social media*, 243–248. ACM.

Kitcher, P., 1995. Contrasting Conceptions of Social Epistemology. In F. Schmitt, ed., *Socializing Epistemology: The Social Dimensions of Knowledge*, 111–134. Rowman and Littlefield.

Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, E.H., Makse, H.A., 2010. Identifying Influential Spreaders in Complex Networks. *Nature Physics*, 6(888–893).

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A., 2014. Multilayer networks. *Journal of complex networks*, 2(3):203–271.

Klimushkin, M., Obiedkov, S., Roth, C., 2010. Approaches to the selection of relevant concepts in the case of noisy data. In L. Kwuida, B. Sertkaya, eds., *Proc. 8th Intl. Conf. Formal Concept Analysis, LNCS/LNAI*, vol. 5986, 255–266. Springer.

Knorr-Cetina, K., 1999. *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Korff, V.P., Oberg, A., Powell, W.W., 2015. Interstitial organizations as conversational bridges. *Bull. Am. Society Inf. Science Technology*, 41(2):34–38.

Koskinen, J., Jones, P., Medeuov, D., Antonyuk, A., Puzyreva, K., Basov, N., 2020. Analyzing Networks of Networks. *arXiv*, 2008(03692).

Koskinen, J.H., Snijders, T.A.B., 2007. Bayesian inference for dynamic social network data. *Journal of Stat Planning and Inference*, 137(12):3930–3938.

Kossinets, G., Watts, D.J., 2006. Empirical Analysis of an Evolving Social Network. *Science*, 311:88–90.

—, 2009. Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2):405–450.

Krackhardt, D., 1987. Cognitive social structures. *Social networks*, 9(2):109–134.

Krackhardt, D., Carley, K.M., 1998. A PCANS model of structure in organization. In *Proc. International Symposium on Command and Control Research and Technology*, 113–119. Monterray, CA.

Kreuzman, H., 2001. A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science. *Scientometrics*, 51(3):525–539.

Kuhn, T., Perc, M., Helbing, D., 2014. Inheritance patterns in citation networks reveal scientific memes. *Physical Review X*, 4(4):041036.

Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., Upfal, E., 2000. Stochastic models for the web graph. In *Proc 41st FOCS Annual Symposium on Foundations of Computer Science*, 57–65. IEEE.

Kuznetsov, S., 1990. Stability as an estimate of degree of substantiation of hypotheses derived on the basis of operational similarity. *Automatic Documentation and Mathematical Linguistics (Nauch. Tekh. Inf. Ser. 2)*, 24(6):62–75.

Kuznetsov, S., Obiedkov, S., Roth, C., 2007. Reducing the Representation Complexity of Lattice-Based Taxonomies. In U. Priss, S. Polovina, R. Hill, eds., *Conceptual Structures: Knowledge Architectures for Smart Applications*, *LNCS*, vol. 4604, 241–254. Springer.

Lancichinetti, A., Fortunato, S., 2012. Consensus clustering in complex networks. *Scientific reports*, 2(1):1–7.

Larremore, D.B., Clauset, A., Jacobs, A.Z., 2014. Efficiently inferring community structure in bipartite networks. *Physical Review E*, 90(1):012805.

Latapy, M., Viard, T., Magnien, C., 2018. Stream graphs and link streams for the modeling of interactions over time. *Social Network Analysis and Mining*, 8(1):1–29.

Latour, B., Woolgar, S., 1979. *Laboratory Life: The Social Construction of Scientific Facts*. Sage.

Lazega, E., 1992. *The micropolitics of knowledge*. Aldine de Gruyter, New York.

Lazega, E., Jourda, M.T., Mounier, L., Stofer, R., 2008. Catching up with big fish in the big pond? Multi-level network analysis through linked design. *Social Networks*, 30(2):159–176.

Lazega, E., Pattison, P., 1999. Multiplexity, Generalized Exchange and Cooperation in Organizations. *Social Networks*, 21:67–90.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M., 2009. Computational Social Science. *Science*, 323(5915):721–723.

Le, Q., Mikolov, T., 2014. Distributed representations of sentences and documents. *PMLR*, 32(2):1188–1196.

Leenders, R.T., 1997. Longitudinal behavior of network structure and actor attributes: Modeling interdependence of contagion and selection. In P. Doreian, F.N. Stokman, eds., *Evolution of social networks*, 165–184. Routledge.

Lehmann, J., Gonçalves, B., Ramasco, J., Cattuto, C., 2012. Dynamical Classes of Collective Attention in Twitter. In *Proc. WWW'12 21st Intl. Conf on World Wide Web*, 251–260. ACM.

Lenclud, G., 1998. La culture s'attrape-t-elle ? *Communications*, 66(1):165–183.

Lerique, S., Roth, C., 2018. The Semantic Drift of Quotations in Blogspace: A Case Study in Short-Term Cultural Evolution. *Cognitive Science*, 42(1):188–219.

Leskovec, J., Adamic, L.A., Huberman, B.A., 2007. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5.

Leskovec, J., Backstrom, L., Kleinberg, J., 2009. Meme-tracking and the dynamics of the news cycle. In *Proc. SIGKDD'09 15th Intl Conf Knowledge Discovery and Data Mining*, 497–506. ACM.

Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., Ghahramani, Z., 2010a. Kronecker Graphs: An approach to modeling networks. *JMLR*, 11:985.

Leskovec, J., Horvitz, E., 2008. Planetary-scale views on a large instant-messaging network. In *Proc 17th WWW Intl Conf World Wide Web*, 915–924. ACM.

Leskovec, J., Huttenlocher, D., Kleinberg, J., 2010b. Governance in social media: a case study of the Wikipedia promotion process. In *Proc ICWSM 4th Intl Conf Weblogs Social Media*, 98–105. AAAI.

Leskovec, J., Kleinberg, J., Faloutsos, C., 2005. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proc KDD 11 Knowledge Disc. Data Mining*, 177–187. ACM.

Lev-On, A., Manin, B., 2009. Happy Accidents: Deliberation and Online Exposure to Opposing Views. In T. Davies, S.P. Gangadharan, eds., *Online Deliberation: Design, Research, and Practice*, chap. 7, 105–122. University of Chicago Press.

Lewis, K., Gonzalez, M., Kaufman, J., 2012. Social selection and peer influence in an online social network. *PNAS*, 109(1):68–72.

Leydesdorff, L., 1997. Why Words and Co-words Cannot Map the Development of the Sciences. *JASIST*, 48(5):418–427.

Leydesdorff, L., Nerghes, A., 2017. Co-word Maps and Topic Modeling: A Comparison Using Small and Medium-Sized Corpora (N < 1,000). *JASIST*, 68(4):1024–1035.

Liang, X., Zhao, J., Dong, L., Xu, K., 2013. Unraveling the origin of exponential law in intra-urban human mobility. *Scientific Reports*, 3(2983):1–7.

Liben-Nowell, D., Kleinberg, J., 2003. The link prediction problem for social networks. In *Proc CIKM 12 Conf Inf Knowledge Management*, 556–559. ACM.

Licklider, J., Taylor, R., 1968. The computer as a communication device. *Science and technology*, 76(2):1–3.

Lietz, H., Wagner, C., Bleier, A., Strohmaier, M., 2014. When Politicians Talk: Assessing Online Conversational Practices of Political Parties on Twitter. In *Proc. ICWSM 8th Intl. Conf. Weblogs and Social Media*, 285–294. AAAI.

Lippi, M., Torroni, P., 2016. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)*, 16(2):10.

Liu, Y., Niculescu-Mizil, A., Gryc, W., 2009. Topic-link LDA: Joint models of topic and author community. In *Proc 26th ICML Intl Conf on Machine Learning*, 665–672. ACM.

Liu, Z., Zheng, V.W., Zhao, Z., Zhu, F., Chang, K.C.C., Wu, M., Ying, J., 2017. Semantic proximity search on heterogeneous graph by proximity embedding. In *Proc 31st Conf. on Artificial Intelligence*, 154–160. AAAI.

Lloyd, A.L., May, R.M., 2001. How Viruses Spread Among Computers and People. *Science*, 292(5520):1316–1317.

Lloyd, L., Kaulgud, P., Skiena, S., 2006. Newspapers vs. blogs: Who gets the scoop? In *Proc AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs (AAAI-CAAW)*, 117–124.

Lorrain, F., White, H.C., 1971. Structural Equivalence of Individuals in Social Networks. *Journal of Mathematical Sociology*, 1(49–80).

Louf, R., Roth, C., Barthelemy, M., 2014. Scaling in Transportation Networks. *PLOS ONE*, 9(7):e102007.

Lü, L., Zhou, T., 2011. Link prediction in complex networks: A survey. *Physica A*, 390:1150–1170.

Luce, R.D., 1950. Connectivity and Generalized Cliques in Sociometric Group Structure. *Psychometrika*, 15(2):169–190.

Lung, R.I., Gaskó, N., Suciu, M.A., 2018. A hypergraph model for representing scientific output. *Scientometrics*, 117(3):1361–1379.

Ma, Z., Sun, A., Cong, G., 2013. On Predicting the Popularity of Newly Emerging Hashtags in Twitter. *JASIST*, 64(7):1399–1410.

Macy, M.W., Kitts, J.A., Flache, A., Benard, S., 2003. Polarization in Dynamic Networks: A Hopfield Model of Emergent Structure. In R. Breiger, K. Carley, P. Pattison, eds., *Dynamic Social Network Modeling and Analysis*, 162–173. National Academies Press, Washington DC.

Mahadevan, P., Krioukov, D., Fall, K., Vahdat, A., 2006. Systematic Topology Analysis and Generation Using Degree Correlations. *ACM SIGCOMM Computer Communication Review*, 36(4):135–146.

Märtens, M., Kuipers, F., Van Mieghem, P., 2017. Symbolic Regression on Network Properties. In *Proc. EuroGP 2017: Genetic Programming*, *LNCS*, vol. 10196, 131–146. Springer.

Martin-Borregon, D., Aiello, L.M., Grabowicz, P., Jaimes, A., Baeza-Yates, R., 2014. Characterization of online groups along space, time, and social dimensions. *EPJ Data Science*, 3(8).

Mausam, Schmitz, M., Soderland, S., Bart, R., Etzioni, O., 2012. Open language learning for information extraction. In *Proc Joint Conf on empirical methods in natural language processing and computational natural language learning*, 523–534.

Mazieres, A., Roth, C., 2018. Large-scale diversity estimation through surname origin inference. *Bulletin of Sociological Methodology*, 139(1):59–73.

McCain, K.W., 1986. Cocited author mapping as a valid representation of intellectual structure. *JASIST*, 37(3):111–122.

McCallum, A., Corradda-Emmanuel, A., Wang, X., 2005. Topic and role discovery in social networks. In *Proc IJCAI'05 Intl Joint Conf on Artificial Intelligence*, 786–791. ACM.

McGlohon, M., Leskovec, J., Faloutsos, C., Hurst, M., Glance, N., 2007. Finding Patterns in Blog Shapes and Blog Evolution. In *Proc. ICWSM Intl Conf Weblogs Social Media*, 18.

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444.

Medeuov, D., Roth, C., Puzyreva, K., Basov, N., 2021. Concept-Centered Comparison of Semantic Networks. In *Complex Networks & Their Applications IX, SCI*, vol. 943, 330–341. Springer.

Mei, Q., Cai, D., Zhang, D., Zhai, C., 2008. Topic modeling with network regularization. In *Proc 17th WWW Intl Conf World Wide Web*, 101–110. ACM.

Melamed, D., 2014. Community structures in bipartite networks: A dual-projection approach. *PLoS ONE*, 9(5):e97823.

Menczer, F., 2004. Evolution of document networks. *PNAS*, 101(S1):5261–5265.

Menezes, T., 2011. Evolutionary modeling of a blog network. In *Proc. CEC'2011 Congress on Evolutionary Computation*, 909–916. IEEE.

Menezes, T., Gargiulo, F., Roth, C., Hamberger, K., 2016. New simulation techniques in kinship network analysis. *Structure and Dynamics*, 9(2).

Menezes, T., Roth, C., 2013. Automatic Discovery of Agent-Based Models: An Application to Social Anthropology. *Advances in Complex Systems*, 16(7):1350027.

—, 2014. Symbolic regression of generative network models. *Scientific Reports*, 4(6284).

—, 2017. Natural Scales in Geographical Patterns. *Scientific Reports*, 7(45823).

—, 2019a. Automatic discovery of families of network generative processes. In F. Ghanbarnejad, R. Roy, F. Karimi, J.C. Delvenne, B. Mitra, eds., *Dynamics on and of Complex Networks*, vol. III: Machine Learning and Stat. Physics, 83–111. Springer.

—, 2019b. Semantic Hypergraphs. *arXiv*, 1908.10784.

Menezes, T., Roth, C., Cointet, J.P., 2010a. Finding the semantic-level precursors on a blog network. *Int J Soc Comp Cyber-Physical Sys*, 1(2):115–134.

—, 2010b. Precursors and Laggards: An Analysis of Semantic Temporal Relationships on a Blog Network. In *Proc. 2nd SocialCom Intl. Conf. Social Computing*, 120–127. IEEE.

Mesoudi, A., Whiten, A., Laland, K.N., 2004. Perspective: Is Human Cultural Evolution Darwinian? Evidence Reviewed from the Perspective of the Origin of Species. *Evolution*, 58(1):1–11.

Mihalcea, R., Tarau, P., 2004. TextRank: Bringing Order into Text. In *EMNLP*, vol. 4, 404–411.

Mika, P., 2005. Social networks and the Semantic Web: the next challenge. *IEEE Intelligent Systems*, 20(1):80—93.

—, 2007. Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, 5:5–15.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013. Distributed representations of words and phrases and their compositionality. *NIPS Adv. in Neur. Inf. Proc. Sys.*, 26:3111–3119.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenhtat, I., Sheffer, M., Alon, U., 2004. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542.

Mische, A., Pattison, P., 2000. Composing a civic arena: Publics, projects, and social settings. *Poetics*, 163–194.

Mitra, B., Tabourier, L., Roth, C., 2012. Intrinsically Dynamic Network Communities. *Computer Networks*, 56(3):1041–1053.

Mitzlaff, F., Atzmueller, M., Stumme, G., Hotho, A., 2013. Semantics of user interaction in social media. In *Complex Networks IV*, 13–25. Springer.

Mohr, J.W., 1994. Soldiers, mothers, tramps and others: Discourse roles in the 1907 New York City charity directory. *Poetics*, 22(4):327–357.

Mohr, J.W., Duquenne, V., 1997. The Duality of Culture and Practice: Poverty Relief in New York City, 1888-1917. *Theory and Society*, 26(2/3):305–356.

Möller, J., Trilling, D., Helberger, N., van Es, B., 2018. Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, 21(7):959–977.

Molloy, M., Reed, B., 1995. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 161(6):161–179.

Moody, J., 2001. Peer influence groups: identifying dense clusters in large networks. *Social Networks*, 23:261–283.

Moody, J., White, D.R., 2003. Structural cohesion and embeddedness: a hierarchical conception of social groups. *Am. Sociological Review*, 68(103–127).

Morales, G.D.F., Monti, C., Starnini, M., 2021. No echo in the chambers of political interactions on Reddit. *Scientific Reports*, 11(1):1–12.

Morris, S., 2000. Contagion. *Review of Economic Studies*, 67(1):57–78.

Moussaïd, M., Brighton, H., Gaissmaier, W., 2015. The amplification of risk in experimental diffusion chains. *PNAS*, 112(18):5631–5636.

Mucha, P., Richardson, T., Macon, K., Porter, M., Onnela, J.P., 2010. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878.

Müller-Birn, C., Dobusch, L., Herbsleb, J., 2013. Work-to-rule: the emergence of algorithmic governance in Wikipedia. In *Proc. 6th Intl Conf Communities and Technologies C&T'13*, 80–89. ACM.

Mullins, N.C., 1972. The Development of a Scientific Specialty: The Phage Group and the Origins of Molecular Biology. *Minerva*, 10(1):51–82.

Munson, S.A., Resnick, P., 2010. Presenting Diverse Political Opinions: How and How Much. In *Proc CHI Human Factors Comp Sys*, 1457–1466. ACM.

Murthy, D., 2008. Digital Ethnography: An Examination of the Use of New Technologies for Social Research. *Sociology*, 42(5):837–855.

Nadeau, D., Sekine, S., 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

Nerghes, A., Lee, J.S., Groenewegen, P., Hellsten, I., 2015. Mapping discursive dynamics of the financial crisis: a structural perspective of concept roles in semantic networks. *Computational Social Networks*, 2(1):1–29.

Newman, M.E.J., 2001. Scientific collaboration networks. I. Network construction and fundamental results, and II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64:016131 & 016132.

—, 2002. Spread of Epidemic Disease on Networks. *Physical Review E*, 66:016128.

—, 2004. Detecting community structure in networks. *European Physical Journal B*, 38:321–330.

—, 2006. Modularity and community structure in networks. *PNAS*, 103(23):8577–8582.

Newman, M.E.J., Strogatz, S., Watts, D., 2001. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64:026118.

Newman, M.E.J., Strogatz, S.H., Watts, D.J., 2002. Random graphs models of social networks. *PNAS*, 99(S1):2566–2572.

Nguyen, T.T., Hui, P.M., Harper, F.M., Terveen, L., Konstan, J.A., 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proc. WWW'14 23rd Intl. Conf. on World Wide Web*, 677–686. ACM.

Niederer, S., van Dijck, J., 2010. Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society*, 12(8):1368–1387.

Nikolov, D., Oliveira, D.F., Flammini, A., Menczer, F., 2015. Measuring online social bubbles. *PeerJ Computer Science*, 1:e38.

Onnela, J.P., Fenn, D.J., Reid, S., Porter, M.A., Mucha, P.J., Fricker, M.D., Jones, N.S., 2012. Taxonomies of networks from community structure. *Physical Review E*, 86:036104.

Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.L., 2007. Structure and tie strengths in mobile communication networks. *PNAS*, 104:7332–7336.

Pachucki, M.A., Breiger, R.L., 2010. Beyond Relationality in Social Networks and Culture. *Annual Review of Sociology*, 36:205–244.

Palla, G., Barabási, A.L., Vicsek, T., 2007. Quantifying social group evolution. *Nature*, 446:664–667.

Pang, B., Lee, L., et al., 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

Papadopoulos, F., Kitsak, M., Serrano, M.A., Boguná, M., Krioukov, D., 2012. Popularity versus similarity in growing networks. *Nature*, 489:537–540.

Pastor-Satorras, R., Vespignani, A., 2001. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, 86(14):3200–3203.

Pathak, N., DeLong, C., Banerjee, A., Erickson, K., 2008. Social topic models for community extraction. In *Proc. 2nd SNA-KDD Workshop*.

Payette, N., 2012. Agent-based models of science. In *Models of science dynamics*, 127–157. Springer.

Pearce, W., Holmberg, K., Hellsten, I., Nerlich, B., 2014. Climate change on Twitter: Topics, Communities and Conversations about the 2013 IPCC Working Group 1 report. *PLOS ONE*, 9(4).

Peixoto, T.P., 2018. Nonparametric weighted stochastic block models. *Physical Review E*, 97(1):012306.

Perra, N., Gonçalves, B., Pastor-Satorras, R., Vespignani, A., 2012. Activity driven modeling of time varying networks. *Scientific Reports*, 2(469).

Poelmans, J., Ignatov, D.I., Kuznetsov, S.O., Dedene, G., 2013. Formal concept analysis in knowledge processing: A survey on applications. *Expert Systems with Applications*, 40(16):6538–6560.

Powell, W.W., White, D.R., Koput, K.W., Owen-Smith, J., 2005. Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*, 110(4):1132–1205.

Pujol, J.M., Flache, A., Delgado, J., Sangüesa, R., 2005. How Can Social Networks Ever Become Complex? Modelling the Emergence of Complex Networks from Local Social Exchanges. *JASSS*, 8(4).

Puschmann, C., 2019. Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism*, 7(6):824–843.

Raimbault, B., Cointet, J.P., Joly, P.B., 2016. Mapping the emergence of synthetic biology. *PLOS ONE*, 11(9):e0161522.

Ramos-Villagrasa, P.J., Marques-Quinteiro, P., Navarro, J., Rico, R., 2018. Teams as complex adaptive systems: Reviewing 17 years of research. *Small Group Research*, 49(2):135–176.

Rao, A., Jana, R., Bandyopadhyay, S., 1996. A Markov chain Monte Carlo method for generating random (0, 1)-matrices with given marginals. *Sankhya: The Indian Journal of Statistics, Series A*, 225–242.

Rheingold, H., 1993. *Virtual Communities: Homesteading on the Electronic Frontier*. MIT Press.

Ritter, A., Clark, S., Etzioni, O., et al., 2011. Named entity recognition in tweets: an experimental study. In *Proc. EMNLP Conf Empirical Methods in Natural Language Processing*, 1524–1534. ACL.

Robertson, T.S., 1967. The process of innovation and the diffusion of innovation. *J. Marketing*, 31(1):14.

Robins, G., Pattison, P., Kalish, Y., Lusher, D., 2007. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191.

Rogers, E.M., 1976. New product adoption and diffusion. *Journal of Consumer Research*, 2(4):290–301.

Romero, D.M., Meeder, B., Kleinberg, J., 2011. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. In *Proc. WWW'11 Intl. Conf. World Wide Web*, 695–704. ACM.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smyth, P., 2004. The Author-Topic Model for Authors and Documents. In *Proc UAI'04 20th Conf Uncertainty in Artificial Intelligence*, 487–494. AUAI Press.

Rossetti, G., Cazabet, R., 2018. Community discovery in dynamic networks: a survey. *ACM Computing Surveys (CSUR)*, 51(2):1–37.

Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. *PNAS*, 105(4):1118–1123.

—, 2010. Mapping change in large networks. *PLOS ONE*, 5(1):e8694.

Roth, C., 2005. Generalized Preferential Attachment: Towards Realistic Socio-Semantic Network Models. In *Proc ISWC 4 Intl Semantic Web Conf, Workshop Sem Net Analysis*, *CEUR-WS*, vol. 171, 29–42.

—, 2006a. Co-evolution in Epistemic Networks – Reconstructing Social Complex Systems. *Structure and Dynamics*, 1(3):2–161.

—, 2006b. Compact, evolving community taxonomies using concept lattices. In P. Hitzler, H. Schaerfe, P. Ohrstrom, eds., *Contributions to 14th ICCS Intl Conf on Conceptual Structures*, 172–187. Aalborg University Press.

—, 2007a. Empiricism for descriptive social network models. *Physica A*, 378(1):53–58.

—, 2007b. La reconstruction en sciences sociales: le cas des réseaux de savoirs. *Nouvelles Perspectives en Sciences Sociales*, 2(2):59–101.

—, 2007c. Morphogenesis of epistemic networks: a case study. In F. Amblard, ed., *Proc. 4th ESSA Conf Eur Soc Simulation Association*, 575–588.

—, 2007d. Viable Wikis – Struggle for Life in the Wikisphere. In A. Désilets, R. Biddle, eds., *WikiSym 3rd Intl Symposium on Wikis*, 119–124. ACM.

—, 2008a. Co-évolution des auteurs et des concepts dans les réseaux épistémiques: le cas de la communauté "zebrafish". *Revue Française de Sociologie*, 48(2):333–367.

—, 2008b. Réseaux épistémiques: formaliser la cognition distribuée. *Sociologie du Travail*, 50:353–371.

—, 2009. Reconstruction Failure: Questioning Level Design. In F. Squazzoni, ed., *Epistemological Aspects of Computer Simulation in the Social Sciences*, *LNCS*, vol. 5466, 89–98. Springer.

—, 2010. Communautés, analyse structurale et réseaux socio-sémantiques. In I. Sainsaulieu, M. Salzbrunn, L. Amiotte-Suchet, eds., *Faire communauté en société – Dynamique des appartenances collectives*, 113–128. PUR.

—, 2013. Socio-Semantic Frameworks. *Advances in Complex Systems*, 16(04n05):1350013.

—, 2017. Knowledge Communities and Socio-Cognitive Taxonomies. In R. Missaoui, S.O. Kuznetsov, S.A. Obiedkov, eds., *Formal Concept Analysis of Social Networks*, LNSN, 1–18. Springer.

—, 2019a. Algorithmic Distortion of Informational Landscapes. *Intellectica*, 70(1):97–118.

—, 2019b. Digital, digitized, and numerical humanities. *Digital Scholarship in the Humanities*, 34(3):616–632.

Roth, C., Basov, N., 2020. The socio-semantic space of John Mohr. *Poetics*, 101437.

Roth, C., Bourgine, P., 2003. Binding Social and Cultural Networks: a Model. *arXiv*, 0309035.

—, 2005. Epistemic Communities: Description and Hierarchic Categorization. *Mathematical Population Studies*, 12(2):107–130.

—, 2006. Lattice-based dynamic and overlapping taxonomies: The case of epistemic communities. *Scientometrics*, 69(2):429–447.

Roth, C., Cointet, J.P., 2010. Social and Semantic Coevolution in Knowledge Networks. *Social Networks*, 32(1):16–29.

Roth, C., Gargiulo, F., Bringé, A., Hamberger, K., 2013. Random alliance networks. *Social Networks*, 35(3):394–405.

Roth, C., Hellsten, I., to appear. Socio-semantic hierarchy of an online conversation space — The case of Twitter users discussing the #IPCC reports.

Roth, C., Kang, S.M., Batty, M., Barthélémy, M., 2011a. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLOS ONE*, 6(1):e15923.

—, 2012a. A long-time limit for world subway networks. *J. Royal Soc. Interface*, 9(75):2540–2550.

Roth, C., Mazières, A., Menezes, T., 2020. Tubes and bubbles – Topological confinement of YouTube recommendations. *PLOS ONE*, 15(4):e0231703.

Roth, C., Obiedkov, S., Kourie, D.G., 2006. Towards Concise Representation for Taxonomies of Epistemic Communities. In *Proc. CLA 4th Intl Conf on Concept Lattices and Applications*, *LNCS*, vol. 4923, 240–255. Springer.

—, 2008a. On Succinct Representation of Knowledge Community Taxonomies with Formal Concept Analysis. *Intl Journal of Foundations of Computer Science (IJFCS)*, 19(2):383–404.

Roth, C., Taraborelli, D., Gilbert, N., 2008b. Measuring wiki viability: an empirical assessment of the social dynamics of a large sample of wikis. In *Proc WikiSym 4th Intl Symp Wikis*, 27, 1–5. ACM.

—, 2008c. Viabilité des wikis. *Réseaux*, 152:205–240.

—, 2011b. Symposium on "Collective representations of quality". *Mind and Society*, 10(2):165–168.

Roth, C., Wu, J., Lozano, S., 2012b. Assessing impact and quality from local dynamics of citation networks. *Journal of Informetrics*, 6(1):111–120.

Rowe, M., Stankovic, M., Alani, H., 2012. Who Will Follow Whom? Exploiting Semantics for Link Prediction in Attention-Information Networks. In P. Cudré-Mauroux, J. Heflin et al., eds., *Proc. ISWC'12 11th Intl Semantic Web Conference, Part I, LNCS*, vol. 7649, 476–491. Springer.

Ruef, M., 2002. A structural event approach to the analysis of group composition. *Social Networks*, 24:135–160.

Ruiz, P., Plancq, C., Poibeau, T., 2016. More than Word Cooccurrence: Exploring Support and Opposition in International Climate Negotiations with Semantic Parsing. In *LREC 10th Language Resources and Evaluation Conference*, 1902–1907.

Ruppert, E., Law, J., Savage, M., 2013. Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society*, 30(4):22–46.

Sachan, M., Contractor, D., Faruquie, T.A., Subramaniam, L.V., 2012. Using content and interactions for discovering communities in social networks. In *Proc. WWW'12 21st Intl. Conf on World Wide Web*, 331–340. ACM.

Sack, W., 2003. Conversation map: a content-based Usenet newsgroup browser. In *From Usenet to CoWebs*, 92–109. Springer.

Saint-Charles, J., Mongeau, P., 2018. Social influence and discourse similarity networks in workgroups. *Social Networks*, 52:228–237.

Salganik, M.J., Dodds, P.S., Watts, D.J., 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311:854–6.

Salton, G., Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523.

Salton, G., Wong, A., Yang, C.S., 1975. Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11):613–620.

Sarkar, P., Chakrabarti, D., Jordan, M., 2014. Nonparametric link prediction in large scale dynamic networks. *Electronic Journal of Statistics*, 8(2):2022–2065.

Sasahara, K., Chen, W., Peng, H., Ciampaglia, G.L., Flammini, A., Menczer, F., 2020. Social influence and unfollowing accelerate the emergence of echo chambers. *J Computational Social Science*.

Sawyer, R.K., 2001. Emergence in sociology: Contemporary philosophy of mind and some implications for sociological theory. *Am. J. Soc.*, 107(3):551–585.

Šćepanović, S., Mishkovski, I., Gonçalves, B., Nguyen, T.H., Hui, P., 2017. Semantic homophily in online communication: evidence from twitter. *Online Social Networks and Media*, 2:1–18.

Schifanella, R., Barrat, A., Cattuto, C., Markines, B., Menczer, F., 2010. Folks in folksonomies: Social link prediction from shared metadata. In *Proc WSDM Web Search Data Mining*, 271–280. ACM.

Schmidt, M., Lipson, H., 2009. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85.

Sewell, W.H.J., 1992. A theory of structure: Duality, agency, and transformation. *American journal of sociology*, 98(1):1–29.

Shahaf, D., Guestrin, C., Horvitz, E., 2012. Trains of thought: Generating information maps. In *Proc. WWW'12 21st Intl. Conf on World Wide Web*, 899–908. ACM.

Shahaf, D., Yang, J., Suen, C., Jacobs, J., Wang, H., Leskovec, J., 2013. Information Cartography: Creating Zoomable, Large-Scale Maps of Information. In *Proc 19th SIGKDD Intl Conf on Knowledge Discovery and Data mining*, 1097–1105. ACM.

Shalizi, C.R., Thomas, A.C., 2011. Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research*, 40(2):211–239.

Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.C., Flammini, A., Menczer, F., 2018. The spread of low-credibility content by social bots. *Nature Communications*, 9(1):4787.

Sharma, A., Srivastava, J., Chandra, A., 2014. Predicting multi-actor collaborations using hypergraphs. *arXiv*, 1401.6404.

Shoham, Y., Leyton-Brown, K., 2008. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*, chap. 13, Logics of Knowledge and Belief, 393–420. Cambridge University Press.

Sim, Y., Acree, B.D., Gross, J.H., Smith, N.A., 2013. Measuring ideological proportions in political speeches. In *Proc EMNLP Conf on Empirical Methods in Nat Lang Proc*, 91–101. ACL.

Simmel, G., 1898. The Persistence of Social Groups. *American Journal of Sociology*, 3(5):662.

Simmons, M.P., Adamic, L.A., Adar, E., 2011. Memes Online: Extracted, Substracted, Injected and Recollected. In *Proc. 5th ICWSM Intl Conf Weblogs and Social Media*, 353–360. AAAI.

Singhal, A., 2001. Modern information retrieval: A brief overview. *IEEE Data Engineering Bulletin*, 24(4):35–43.

Smaldino, P.E., Janssen, M.A., Hillis, V., Bednar, J., 2017. Adoption as a social marker: Innovation diffusion with outgroup aversion. *Journal of Mathematical Sociology*, 41(1):26–45.

Small, H., Griffith, B.C., 1974. The structure of scientific literatures I: Identifying and graphing specialties. *Science studies*, 4(1):17–40.

Snijders, T.A.B., 2001. The Statistical Evaluation of Social Networks Dynamics. *Sociological Methodology*, 31(1):361–395.

Snijders, T.A.B., Steglich, C., Schweinberger, M., 2007. Modeling the co-evolution of networks and behavior. In K. van Montfort, J. Oud, A. Satorra, eds., *Longitudinal models in the behavioral and related sciences*, 41–71. Lawrence Erlbaum.

Sobolevsky, S., Szell, M., Campari, R., Couronné, T., Smoreda, Z., Ratti, C., 2013. Delineating geographical regions with networks of human interactions in an extensive set of countries. *PLOS ONE*, 8(12):e81707.

Sperber, D., 1996. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell Publishers.

Srivastava, A.N., Sahami, M., 2009. *Text mining: Classification, clustering, and applications*. Data Mining and Knowledge Discovery. CRC Press.

Stokols, D., Hall, K.L., Taylor, B.K., Moser, R.P., 2008. The Science of Team Science. *American Journal of Preventive Medicine*, 35:S78–S89.

Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L., 2002. Computing iceberg concept lattices with TITANIC. *Data and Knowledge Engineering*, 42:189–222.

Sun, X., Kaur, J., Milojević, S., Flammini, A., Menczer, F., 2013. Social dynamics of science. *Scientific reports*, 3(1):1–6.

Sun, Y., Han, J., Zhao, P., Yin, Z., Cheng, H., Wu, T., 2009. RankClus: integrating clustering with ranking for heterogeneous information network analysis. In *Proc. EDBT 12th Intl Conf on Extending Database Technology*, 565–576. ACM.

Sunstein, C.R., 2007. *Republic.com 2.0*. Princeton University Press.

Tabourier, L., Cointet, J.P., Roth, C., 2017. Génération de graphes aléatoires par échanges multiples d'arêtes. *Journal de la Société Française de Statistiques*, 158(2):118—134.

Tabourier, L., Roth, C., Cointet, J.P., 2011. Generating constrained random graphs using multiple edge switches". *ACM Journal of Experimental Algorithmics*, 16(1.7).

Tamine, L., Soulier, L., Ben Jabeur, L., Amblard, F., Hanachi, C., Hubert, G., Roth, C., 2016. Social Media-Based Collaborative Information Access: Analysis of Online Crisis-Related Twitter Conversations. In *Proc. HT'16 27th Intl Conference on Hypertext and Social Media*, 159–168. ACM.

Tancoigne, E., Barbier, M., Cointet, J.P., Richard, G., 2014. The place of agricultural sciences in the literature on ecosystem services. *Ecosystem Services*, 10:35–48.

Taraborelli, D., Roth, C., 2011. Viable web communities: Two case studies. In G. Deffuant, N. Gilbert, eds., *Viability and resilience of complex systems*, 75–105. Springer.

Taramasco, C.A., Cointet, J.P., Roth, C., 2010. Academic team formation as evolving hypergraphs. *Scientometrics*, 85(3):721–774.

Thiemann, C., Theis, F., Grady, D., Brune, R., Brockmann, D., 2010. The structure of borders in a small world. *PLOS ONE*, 5(11):e15422.

Thomas, K., Grier, C., Paxson, V., 2012. Adapting Social Spam Infrastructure for Political Censorship. In *Proc. 5th USENIX Workshop on Large-Scale Exploits and Emergent Threats*, 1–8.

Tilly, C., 1997. Parliamentarization of Popular Contention in Great Britain, 1758-1834. *Theory and Society*, 26(2/3):245–273.

Traag, V.A., Krings, G., Dooren, P.V., 2013. Significant Scales in Community Structure. *Scientific Reports*, 3(2930):1–10.

Traag, V.A., Van Dooren, P., Nesterov, Y., 2011. Narrow scope for resolution-limit-free community detection. *Physical Review E*, 84(1):016114.

Turner, F., 2006. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. University of Chicago Press.

Tyagi, A., Babcock, M., Carley, K.M., Sicker, D.C., 2020. Polarizing tweets on climate change. In *Social, Cultural, and Behavioral Modeling*, *LNCS*, vol. 12268, 107–117. Springer.

Uzzi, B., Spiro, J., 2005. Collaboration and Creativity: the Small-World Problem. *American Journal of Sociology*, 111(2):447–504.

Vaisey, S., Lizardo, O., 2010. Can Cultural Worldviews Influence Network Composition? *Social Forces*, 88(4):1595–1618.

Valente, T.W., 1996. Social network thresholds in the diffusion of innovations. *Social Networks*, 18:69.

Van Alstyne, M., Brynjolfsson, E., 1996. Electronic Communities: Global Village or Cyberbalkans? In J. DeGross, S. Jarvenpaa, A. Srinivasan, eds., *Proc 17 ICIS Intl Conf Information Systems*, 80–98.

Van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., 2008. Parsing, semantic networks, and political authority using syntactic analysis to extract semantic relations from Dutch newspaper articles. *Political Analysis*, 16(4):428–446.

Van Atteveldt, W., Sheafer, T., Shenhav, S.R., Fogel-Dror, Y., 2017. Clause analysis: using syntactic information to automatically extract source, subject, and predicate from texts with an application to the 2008–2009 Gaza War. *Political Analysis*, 25(2):207–222.

Vázquez, A., 2003. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67:056104.

Viard, J., Latapy, M., Magnien, C., 2016. Computing maximal cliques in link streams. *Theoretical Computer Science*, 609:245–252.

Vilhena, D.A., Foster, J.G., Rosvall, M., West, J.D., Evans, J., Bergstrom, C.T., 2014. Finding Cultural Holes: How Structure and Culture Diverge in Networks of Scholarly Communication. *Sociological Science*, 1:221–238.

Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.

Vossen, P., Agerri, R., Aldabe, I., Cybulska, A., van Erp, M., Fokkens, A., Laparra, E., Minard, A.L., Aprosio, A.P., Rigau, G., Rospocher, M., Segers, R., 2016. NewsReader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. *Knowledge-Based Systems*, 110:60–85.

Waltman, L., van Eck, N.J., 2012. The inconsistency of the h-index. *JASIST*, 63(2):406–415.

Wang, D., Cui, P., Zhu, W., 2016. Structural deep network embedding. In *Proc. 22nd SIGKDD Intl Conf on Knowledge discovery and data mining*, 1225–1234. ACM.

Wang, F.Y., Carley, K.M., Zeng, D., Mao, W., 2007. Social Computing: From Social Informatics to Social Intelligence. *Intelligent Systems*, 22(2):79–83.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S., 2012. A system for real-time twitter sentiment analysis of 2012 US presidential election cycle. In *Proc. 50th ACL Meeting*, 115–120.

Wang, P., Robins, G., Pattison, P., Lazega, E., 2013. Exponential random graph models for multilevel networks. *Social Networks*, 35:96–115.

Wang, S., Groth, P., 2010. Measuring the Dynamic Bi-directional Influence between Content and Social Networks. In *Proc. ISWC 9th Intl Semantic Web Conf*, vol. 6496, 814–829. LNCS, Springer.

Wang, X., Sirianni, A.D., Tang, S., Zheng, Z., Fu, F., 2020. Public discourse and social network echo chambers driven by socio-cognitive biases. *Physical Review X*, 10(4):041042.

Wasserman, S., 1980. Analyzing Social Networks As Stochastic Processes. *Journal of the American Statistical Association*, 75(370):280–294.

Wasserman, S., Faust, K., 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Wasserman, S., Pattison, P., 1996. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika*, 61(3):401–425.

Wattenberg, M., 2006. Visual Exploration of Multivariate Graphs. In *Proc SIGCHI Conf Human Factors in Computing Systems*, 811–819. ACM.

Watts, D.J., Dodds, P.S., 2007. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34(4):441–458.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442.

Weingart, P., 2005. Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62(1):117–131.

Weisberg, M., Muldoon, R., 2009. Epistemic Landscapes and the Division of Cognitive Labor. *Philosophy of Science*, 76(2):225–252.

Wellman, B., 2001. Computer Networks as Social Networks. *Science*, 293:2031–2034.

Weng, L., Flammini, A., Vespignani, A., Menczer, F., 2012. Competition among memes in a world with limited attention. *Scientific Reports*, 2:335.

Weng, L., Menczer, F., Ahn, Y.Y., 2013. Virality Prediction and Community Structure in Social Networks. *Scientific Reports*, 3:2522.

White, H.C., 1992. *Identity and Control: A Structural Theory of Social Action*. Princeton University Press.

White, H.C., Boorman, S.A., Breiger, R.L., 1976. Social-structure from multiple networks. I: Blockmodels of roles and positions. *American Journal of Sociology*, 81:730–780.

Wilkerson, J., Casas, A., 2017. Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, 20:529–544.

Wilkinson, D.M., Huberman, B.A., 2007. Cooperation and quality in wikipedia. In *WikiSym 3rd Intl Symposium on Wikis*, 157–164.

Wilson, S.M., Peterson, L.C., 2002. The Anthropology of Online Communities. *Annual Review of Anthropology*, 31:449–467.

Wu, F., Huberman, B.A., Adamic, L.A., Tyler, J.R., 2004. Information Flow in Social Groups. *Physica A*, 337:327–335.

Xie, J., Kelley, S., Szymanski, B.K., 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys*, 45(4):1–35.

Yanardag, P., Vishwanathan, S., 2015. Deep graph kernels. In *Proc 21 SIGKDD Intl Conf knowledge discovery and data mining*, 1365–1374. ACM.

Yang, J., McAuley, J., Leskovec, J., 2013. Community detection in networks with node attributes. In *Proc 13 Intl Conf Data mining*, 1151–1156. IEEE.

Yang, Y., Lichtenwalter, R.N., Chawla, N.V., 2015. Evaluating link prediction methods. *Knowledge and Information Systems*, 45(3):751–782.

Yaveroglu, O.N., Malod-Dognin, N., Devis, D., Levnajic, Z., Janjic, V., Karapandza, R., Stojmirovic, A., Przulj, N., 2014. Revealing the Hidden Language of Complex Networks. *Scientific Reports*, 4(4547).

Yook, S.H., Jeong, H., Barabási, A.L., 2002. Modeling the Internet's large-scale topology. *PNAS*, 99(21):13382–13386.

Yuan, G., Murukannaiah, P.K., Zhang, Z., Singh, M.P., 2014. Exploiting sentiment homophily for link prediction. In *Proc. RecSys '14 8th Conference on Recommender systems*, 17–24. ACM.

Zacklad, M., Cahier, J.P., Pétard, X., 2003. Du Web cognitivement sémantique au Web socio sémantique, Exigences représentationnelles de la coopération. In *Workshop Web Sémantique & SHS*.

Zheleva, E., Sharara, H., Getoor, L., 2009. Co-evolution of Social and Affiliation Networks. In *Proc SIGKDD 15th Intl Conf on Knowledge Discovery and Data Mining*, 1007–1015. ACM.

Zhou, D., Huang, J., Schölkopf, B., 2006a. Learning with hypergraphs: Clustering, classification, and embedding. *NIPS*, 19:1601–1608.

Zhou, D., Ji, X., Zha, H., Giles, C.L., 2006b. Topic Evolution and Social Interactions: How Authors Effect Research. In *Proc. CIKM 15th Intl Conf Inf Knowledge Management*, 248–257. ACM.

Zhou, D., Manavoglu, E., Li, J., Giles, C.L., Zha, H., 2006c. Probabilistic models for discovering e-communities. In *Proc. 15th WWW Intl Conf on World Wide Web*, 173–182. ACM.

Zhou, Y., Cheng, H., Yu, J.X., 2009. Graph clustering based on structural/attribute similarities. *Proc. VLDB Endowment*, 2(1):718–729.

Zhu, M., Huang, Y., Contractor, N.S., 2013. Motivations for self-assembling into project teams. *Social networks*, 35(2):251–264.

Zuiderveen Borgesius, F.J., Trilling, D., Möller, J., Bodó, B., de Vreese, C.H., Helberger, N., 2016. Should we worry about filter bubbles? *Internet Policy Review*, 5(1).