T

The Nature of Social Structures

Luca Maria Aiello
Nokia Bell Labs, Cambridge, UK

Synonyms

Meaning of social structures; Purpose; Type

Glossary

Pragmatics	A subfield of linguistics that studies the way in which *context* contributes to meaning. It is also concerned with the factors that govern our choice of language in social interaction and with the *effect* of these choices on others. Pragmatics is needed to reach a deeper and more reasonable understanding of human language and, in turn, to grasp the fundamental rules governing social interactions.
Social exchange	The process of transferring nonmaterial resources (e.g., knowledge, affection) between people during social interaction. Individuals contribute to and derive benefits from social structures by means of those exchanges.
Social group	A social aggregation of multiple actors who share a common trait, purpose, or identity and who usually interact with one another. Groups vary greatly in size and type: a small family, a community of fans, or a whole ethnicity can all be considered groups. The boundaries of a group are often defined by its own members; for this reason, an alternative – yet simplistic – definition of group is the collection of people who identify themselves as members of the group.
Social structure	The arrangement of two or more people that emerges from their individual social actions. Social structures are organized hierarchically. On a *macro*scale, people unite in large societies within nations or even across the boundaries of countries; on a *meso*scale, they take part to groups (or communities) with different size and scope; on a *micro*scale, they create bonds with other individuals. The interactions between the hierarchical layers shape norms and behavior of people within a complex social system.
Social tie	Also referred to as *interpersonal* or *dyadic* ties, social ties (or links) are

R. Alhajj, J. Rokne (eds.), *Encyclopedia of Social Network Analysis and Mining*,
DOI 10.1007/978-1-4614-7163-9_110180-1

connections between pairs of people along which any type of social exchange is performed. They can be established and maintained through a variety of means ranging from face-to-face interaction to Internet-mediated tools. They can be rather elusive to quantify because of their inherent dynamic nature and because of the blurred line that in many real-life scenarios divides socially connected individuals from strangers. A set of social ties, together with the set of people connected by those ties, forms a *social network* (or *graph*).

Definition

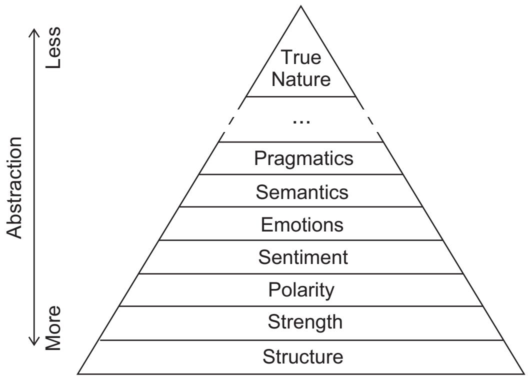
The exploration of the *nature* of social structures is a line of research that aims at quantitatively describing the fundamental social processes that lead to the creation, maintenance, and possibly destruction of the fundamental structural atoms that compose a society: social links and social groups. The rationale behind this pursuit is that, to understand the purpose of social structures as perceived by the agents that are part of them, one needs to go beyond the use of traditional network analysis descriptors such as tie strength or polarity to explore the semantics, pragmatics, and psychology that underlie social exchanges within those structures. This goal can be achieved through the operationalization of sociological theories: High-level social exchange processes, whose existence are indicated by social science literature but are not directly measurable, are quantified through algorithms powered by lower-level, measurable signals. Given the variety and complexity of the facets that define dyadic social relationships in real life, let alone higher-order social structures, this area of research needs to draw concepts from multiple disciplines including network science, sociometrics, social psychology, and computational social science, among others.

Introduction

In social networks, not all ties are created equal. To understand what social links really mean, one needs to handle the intricacies of the context around them. A relationship between a mother and her son has arguably very different characteristics than a typical relationship between two colleagues. In turn, pairs of colleagues who have been knowing each other for the same amount of time might talk about different topics and follow different conversational conventions; one pair can be more focused on work-related conversations, another can be more engaged in everyday chitchatting. The same goes for more complex social structures such as groups. For example, members of a small amateur photography club will most likely have very different interaction dynamics than an equally sized group of school mates.

When modeling a social system with a graph, gaining some level of understanding of what type of interaction the links represent could be crucial. Depending on the goal of the work, the outcome of any analysis that relies on social network data might change radically when the information about the nature of social structures is factored in. As an example, think about two widely studied concepts in network science: graph centrality and information diffusion. The most central actor in a network of acquaintances might not be as central when considering only the ties that are expression of reciprocal trust. Also, the above mentioned photography club would be more permeable to adopt a new photographic technology than any generic group of friends.

The description of social structures can take place at multiple levels of abstraction, as summarized in Fig. 1. At the most basic *structural* level, we can express the existence (or absence) of a social relationship, possibly specifying the directionality of the interaction (e.g., following on Twitter). Network analysis on simple, unweighted graphs operates on this first structural layer. On top of structure, links can be annotated with weights to model the interaction *strength*.



The Nature of Social Structures, Fig. 1 The hierarchy of abstraction layers that can be used to describe social structures, from the most abstract (*bottom*) to the most detailed (*top*)

Frequency, intensity, and volume of interactions are all proxies of the strength of a tie. Extending the model to use negative weights allows for mapping the connection's *polarity* (e.g., friends vs. foes). When textual content about the relationship is available (e.g., text of online-mediated communications), polarity models can be expanded into the analysis of *sentiment* conveyed by the conversation's wording. Sentiment can be a bipolar attribute (i.e., positive, neutral, negative) or a multifaceted dimension (e.g., anger, fear, happiness, anticipation). More in general, the availability of textual data enables the exploration of the linguistic aspect at the level of *syntax* and *semantics*. In particular, topic models can be built to extract concise descriptors of the themes a conversation touches upon. The *pragmatics* layer sits on top of semantics. From the angle of pragmatics, ties can be interpreted as sequences of *communicative acts* that contribute to the incremental definition of the nature of the social relationship between pairs of individuals.

To better understand the meaning of this hierarchical categorization, consider the following stylized example. Imagine a researcher addressing a colleague by saying: *"You are a brilliant scientist."* The social relationship between the two, as exemplified by this anecdotal situation, can be described on all the layers that we have previously defined. First, there exists a binary link between the two actors, which in turn has a specific structural position within the larger social networks in which the two are embedded. The strength of the tie could be roughly estimated by the number of messages or by the number of words exchanged (in the example, one and five, respectively). The message seems to convey a positive sentiment, and it might be categorized under topics such as "research" or "academic relationships." Last, at the level of pragmatics, the message exchange implies that the speaker is expressing admiration and esteem for the alter; spelling that out has the effective power of changing reality, as it contributes to shape the relationship status of those two people, as perceived by both.

The conceptual scheme reported here is nothing but a sketch of a possible hierarchical relationship between layers of abstraction over the true nature of social links and groups. Additional orthogonal dimensions such as space and time can add nuances to this representation. Also, one could argue that there might be several additional layers in between and on top of the ones we have defined. Especially, social science literature over the last century has provided a way more nuanced description of the facets and corners that contribute to draw a more faithful picture of the nature of social structures. Here, we consider this topic mostly from a computational social science perspective, focusing only on the layers that have been investigated using computational methods and especially using online-mediated interaction data.

Surprisingly, despite the exponential growth of online social media penetration has produced an unprecedented volume of multidimensional data about digital social interactions, and the nature of social structures emerging from those interactions has been rarely considered in mainstream research. Online exchanges have been interpreted as indicative of one social process or another (e.g., status giving or trust), often with little systematic justification regarding the relation between observed data and any theoretical concept. Social ties are usually treated as a priori quanta of information, immediately available to the analyst from the graph of online-mediated interactions. On the same note, social groups have been mostly considered as homogeneous entities, overlooking the fact that they may emerge from very different collective processes and from the different motivations of their founders and members.

With the purpose of providing an introduction to this research field, we will next present a non-exhaustive overview of work in computer science, network science, and computational social science research that contributed to bridge gaps between the abstract representation of social structures and their actual nature. For the sake of providing examples of those that might be future research trends in the area, we will then focus on two recent pieces of work that attempted to push the interpretation of social structures to the pragmatics layer and beyond.

Key Points

The branch of computational social science that studies the nature of social structures aims to move past the representation of social relationships as simple links between nodes in a social network and of social groups as mere collections of those nodes. The goal is to extract the first principles that can quantitatively describe, in the most concise way possible, what social structures mean for the people who are part of them.

Historical Background

Social Links

Early work in social networks analysis has focused on social systems modeled as static networks with simple nodes and edges (Wasserman and Faust 1994; Newman 2003). Those abstract models were soon extended to incorporate edge weight (Barrat et al. 2004) to map the concept of *tie strength* on *interaction* networks (Viswanath et al. 2009; Wilson et al. 2009). Tie strength is a concept of crucial importance in explaining the evolution of social networks. One of the most influential publications on this aspect has been authored by sociologist Mark Granovetter (1973), who connected the concept of tie strength with the propensity of people to close social triangles and with how the information flows along social network links. In computer science research, stemming directly from Granovetter's work, Gilbert and Karahalios (2009) used a supervised method to predict the tie strength in Facebook using multiple features. Xiang et al. (2010) addressed the same problem with an unsupervised model instead, using a latent variable model based on some profile features, assuming that the higher the profile similarity the higher the strength of the link. Grabowicz et al. (2012) studied the strength of ties in relation to the communities in the Twitter interaction graphs and identified weak ties as the ones departing from community intermediaries. Besides link strength, some research has been done on the sign of edges in networks with positive and negative links (Kunegis et al. 2009, 2013; Leskovec et al. 2010) (e.g., Slashdot).

Besides network structure, textual records of online conversations have been studied extensively in the context of social media, as they are important drivers of user *engagement* in online communities (Harper et al. 2007). Researchers attempted to identify general predictive models of the main traits of human communication. Research on data from Twitter and email investigated the *conventions* used during conversations (Honeycutt and Herring 2009; Boyd et al. 2010; Kooti et al. 2015) and the evolution of the topics of discussion (Purohit et al. 2014). Backstrom et al. (2013) tried to predict the length of a discussion thread in Facebook using time and content features. Models to reproduce some statistical properties of threads (e.g., size of thread, number of participants) were tested successfully in Twitter and Yahoo Groups (Kumar et al. 2010). Multimodal features of discussion threads can also predict its perceived interestingness (De Choudhury et al. 2009). Correa et al. (2010) conducted interviews to investigate the correlation between psychological indicators, such as emotional stability and openness to new experiences, with propensity to engage online conversations. On a similar note, Celli and Rossi (2012) studied Twitter conversational data, estimated the user emotional stability from the text, and correlated it with the tendency to engage conversations. Some looked at the content of Twitter messages to identify the purpose that users have when communicating with others. For example, Java et al. (2007) manually classified tweets into conversational tweets, messages to share information or report news, and daily chatter.

The development of increasingly accurate algorithms to monitor sentiment and emotions from short text (Gonçalves et al. 2013) has paved the way to analyze the emotional layer of online conversations. Kim et al. (2012) extracted LDA topics in Twitter conversations, used a framework based on the Plutchik's emotion model (Plutchik 1980) to assign emotions to them, and analyzed the transitions between emotions in conversations. They verified that a conversation that conveys a certain sentiment tends do it consistently in the following exchanges ("nice words for nice words").

So far, little attention has been devoted to characterize the type of social links according to sociological dimensions, and recent work on the accommodation of linguistic styles according to power differentials provides an example of the intellectual opportunities now available at the intersection of social theory and conversational data (Danescu-Niculescu-Mizil et al. 2012). Budak et al. identified the presence of high-level domains of interaction in Twitter such as emotional support and information exchange, and studied how those domains relate to the sense of community that is developed by the users. A further step to fill that gap has been done by Bramsen et al. (2011), who have introduced a supervised approach to identify power relationships in social dyads using ad hoc texual features. More recently, more nuanced studies have been done around online conversations, touching upon the concepts of social cohesion and social identity and their implications on group discussion divergence (Purohit et al. 2014), and discussing the social power dynamics they contribute to create (Tchokni et al. 2014).

Research in temporal, dynamic, and multilayer networks are tangentially relevant to the studies investigating the nature of social ties. Influential papers in those areas that provide excellent introductions to the topics have been written by Holme and Saramäki (2012) temporal networks, Gautreau et al. (2009), and Kivela et al. (2014).

Social Groups

Since the very early stages of the social web, the research community has been interested in the definition of the notion of group and of its possible types (Porter 2004), not only for analytical purposes but also in direct application to several tasks, including profiling and recommendation (De Choudhury 2009; Wang et al. 2012). The global structure, evolution, and dynamics of social groups have been investigated over large-scale and heterogeneous datasets. The shape and evolution of groups have been described in computer science literature as very broad phenomena (Mislove et al. 2007; Cox et al. 2011) that are determined by the intrinsic group fitness (Grabowicz and Eguíluz 2012) and on the density of social links connecting their members (Backstrom et al. 2006).

Although the broad variety of group types and their emerging features (starting from their size (Baldassarri et al. 2008)) has motivated some research work to characterize the nature of groups along some of their main measurable dimensions, most of the contributions so far have not established any quantitative framework for their classification. As a consequence, the results obtained in this area are quite scattered.

Due to its open nature, Flickr has been one of the most studied platform to this respect. Early work relied on interviews and user studies to identify the different usage of Flickr groups (Van House 2007), finding five main motivations for users to join groups (memory, identity and narrative, relationships maintenance, self-representation, and self-expression). Alternative classifications based on user studies have been proposed as well (Miller and Edwards 2007; Nov et al. 2010).

Negoescu et al. have contributed to this research area with several studies on Flickr groups. First they have introduced a manual categorization of Flickr groups, partitioning them in *geographical, topical, visual*, and *"catch-all"* groups (Negoescu and Gatica-Perez 2008a). With this categorization in mind, they propose to detect hypergroups (i.e., groups of groups) based on the similarity of their topical focus, extracted with LDA (Negoescu et al. 2009); in contrast, Negi et al. try to find subgroups in large Flickr communities using MoM-LDA on photo tags (Negi and Chaudhury 2012). Groups have been also studied in relation to their membership, with special attention to topicality and to recommendations exchanged between peers (Negoescu and Gatica-Perez 2008b). In more recent work (Negoescu and Gatica-Perez 2010), Negoescu et al. have discussed about how to represent Flickr groups according to the topics and tags in use by their members; according to previous studies (Van House 2007), they identified "real" groups as those motivated by self-expression and relationship maintenance.

Following an earlier conceptual framework (Butler 1999), Cox et al. (2011) attempted to measure the "groupness" of a Flickr group using several metrics including number of members, volume of contributions, length of description, and so on. They propose a classification of groups into topical (focused on a theme), highlighting (to promote photos to a wider public), and geographical (rooted into a specific geolocation); however their classification is ultimately arbitrary and not supported by quantitative results. In partial contrast with previous work (Negoescu and Gatica-Perez 2010), their results also point out that small groups are more important than the big ones to the social activity of the network as they operate at "human scale." The work was subsequently extended (Holmes and Cox 2011) and the categorization was manually refined into four categories, namely location-based, award, learning, and topical groups.

Prieur et al. use PCA on a set of features extracted from Flickr groups to detect the main dimensions that characterize them (Pissard and Prieur 2007; Prieur et al. 2008a, b). They find three main dimensions underlying as many types of groups: social media-use, MySpace-like, and photo stockpiling. The mixture of sociality and topicality of groups is discussed as well.

At a finer scale, social communities can be described in terms of user engagement. From a quantitative perspective, the amount of participation of members in activities related to the group is varied and dependent on group size (Backstrom et al. 2008). Intragroup activity has been characterized in terms of propensity of people to reply to questions of other members (Welser et al. 2007), coherence of discussion topics (Gloor and Zhao 2006), or item-sharing practices (Negoescu and Gatica-Perez 2008a). Modeling inner activity of groups has helped in finding effective strategies to predict future group growth or activity (Kairam et al. 2012), recommend group affiliation, or enhance the search experience on social platforms (Negoescu et al. 2009).

Groups have been studied also in other online platforms. The structure of user interaction patterns in groups extracted from LiveJournal, DBLP, YouTube, Orkut, and Yahoo Groups have been investigated in the past (Spertus et al. 2005; Backstrom et al. 2006, 2008; Mislove et al. 2007). Laine et al. (2011) present an analysis on YouTube groups, highlighting their tendency to both topicality and sociality and on the small-world nature of the interactions inside them.

Besides the analysis of user-created groups, the study of automatically detected groups through community detection algorithms has attracted much interest lately (Fortunato 2010). Detected communities are supposed to represent meaningful aggregations of people where dense or intense social exchanges take place among members (Grabowicz et al. 2012). Nevertheless, even if synthetic methods to verify the quality of clusters have been proposed (Lancichinetti et al. 2008), the question of whether such artificial groups capture some notion of community perceived by the users remains open. If on the one hand the computation of cluster goodness metrics over user-created groups can give useful hints about their structural cohesion (Yang and Leskovec 2015), on the other hand a direct comparison between user-created groups and detected communities is still missing, particularly in terms of the amount of sociality or topical coherence they embed.

Recent work in computational social science attempted to characterize groups in relation to well-established theories from social sciences. The dependency of group activity on group size has been studied in several platforms (Grabowicz et al. 2012; Kairam et al. 2012; Goncalves et al. 2011). The results support from different angles Robin Dunbar's *social brain* theory, which states that human beings can hardly maintain more than 150–200 stable social relationships (Dunbar 1998). Similarity between members has been identified also as an important factor driving the creation of social communities (Tang et al. 2011), particularly given that, to a large extent, users in social networks tend to aggregate following the homophily principle (Aiello et al. 2012). However, similarity is not necessarily the strongest indicator for group activity and longevity, as diversity of content shared between group members is a major factor to keep alive the interest of members (Ludford et al. 2004).

The Nature of Social Links and Social Groups

> Most human pleasures have their roots in social life. [. . .] Much of human suffering as well as much of human happiness has its source in the actions of other human beings. One follows from the other, given the facts of group life, where pairs do not exist in complete isolation from other social relations.

This is how sociologist Peter Blau introduces the discussion about the structure of social associations in his book Exchange and Power in Social Life (Blau 1964), acknowledging the pivotal role of social groups and social links in providing motivations and rewards for people in a social ecosystem. Next, we present two pieces of computational social science research that have been conducted to learn more about the nature of such social structures from large-scale online data.

Identity and Bonding in Groups

Understanding *why* groups form is not as easy as measuring their boundaries, density, or topical focus. Because the notion of group itself hides an enormous variety of concepts representing as many group types, it is very difficult to define what the purpose of a group is.

One simple and well-established interpretation of the meaning of groups is based on the notion of *social identity*. Social psychologist Henri Tajfel, one of the pioneers in the field, describes social identity as the part of an individual's self-concept deriving from the membership of a social group, together with the emotional valuation that the membership may imply (Tajfel 1981). Supporters of a political party, people who suffered from the same illness, members of a fan club, and people interested in the same hobby are all examples of groups that are defined by a common identity.

Social identity is not the only driver of the creation of groups, though. Personal social relations between members, other than shared identity, constitute the backbone that allows some group to form and evolve. Family members stay together and care for each other not only because they share the same family name but mostly because of the strong bonding between individual members. A group of old friends can be very cohesive mostly because of the friendship ties that connect all them.

Psychologist Deborah Prentice is one of the most influential scientists who studied the mechanisms that lead to the creation of a group based on shared identity or presence of strong social bonds. In her formalization of the *common identity and common bond* theory (Prentice et al. 1994), she observes that groups can be categorized as either *identity-based* or *bond-based*, depending on the prevalent motivation that leads people to join the group. The two group types have usually distinct and well-recognizable traits. Identity-based attachment holds when people join a group based on their interest in the community as a whole or in a well-defined common theme shared by all of the members. People whose participation is due to identity-based attachment may not directly engage with anyone and might even participate anonymously. Conversely, bond-based attachment is driven by personal social relations with other specific members, and thus the main theme of the group may be disregarded. The two processes result in two different group types; for simplicity of exposition, we refer to those two categories as *topical* and *social* groups, respectively.

According to the theory, the group type is closely connected to its level of *reciprocity* and its *topics* of discussion. Members of social groups tend to have reciprocal interactions with other members, whereas interactions in topical groups are generally not directly reciprocated. In addition, topics of discussion tend to vary drastically and cover multiple subjects in social groups, while in topical groups discussions tend to be related to the group theme and cover specific topical areas only.

The common identity and common bond theory is an example of how a description of social structures can go beyond its structural and semantic definitions and provide a partial explanation of their *functional nature*. Independent data-driven studies have investigated social and thematic components of groups, but always in separation (Cox et al. 2011). Preliminary insights on the interweavement between such dimensions have been given in exploratory work on Flickr, where signals of correlation between social density and tag dispersion in groups is shown (Prieur et al. 2008b) and where two different clusters emerge naturally when plotting the groups' size against the number of internal links (Baldassarri et al. 2008). Grabowicz et al. (2013) were the first to provide a simple methodology to operationalize the theory by combining measurements of both topical and social aspects of groups. Next, we summarize the key corners of their work.

Let E_g be the number of all the directional social links between the members of a group g, $E_g^{\leftrightarrow} \leq E_g$ the number of those links that are reciprocated, and $T(g)$ the bag of words used by the members of the group to communicate to each other. Group reciprocity increases as members interact with one another in a bidirectional fashion. Numerically, the *reciprocity* of a group g is defined as:

$$r_g = \frac{\frac{E_g^{\leftrightarrow}}{2}}{\frac{E_g^{\leftrightarrow}}{2} + \left(E_g - E_g^{\leftrightarrow}\right)} = \frac{E_g^{\leftrightarrow}}{2E_g + E_g^{\leftrightarrow}}, \tag{1}$$

which, in simple words, is the fraction of connected group member pairs who have a reciprocal interaction. To obtain a score comparable across groups, the reciprocity score is then normalized by the average reciprocity value $\langle r_g \rangle$ over all groups in the social system considered:

$$\hat{r}_g = \frac{r_g}{\langle r_g \rangle}. \tag{2}$$

According to the theory, the larger the intrareciprocity the higher the probability that the group is social. The bag of terms $T(g)$ associated with a group indicates the topical diversity of the group, instead. The group entropy is measured as:

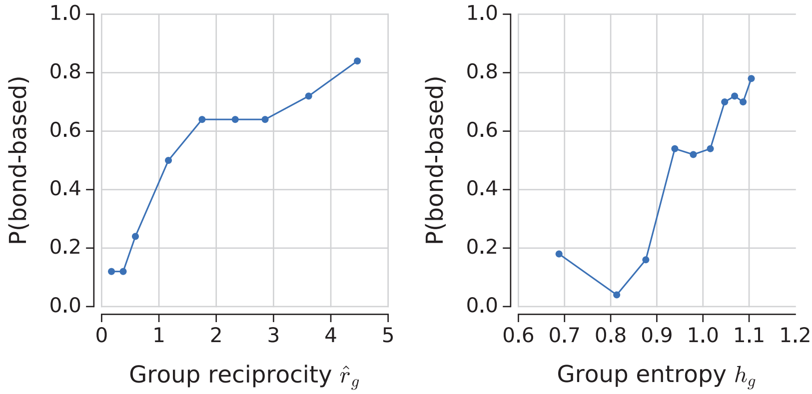
$$H(g) = - \sum_{t \in T(g)} p(t) \cdot \log_2 p(t), \tag{3}$$

where $p(t)$ is the probability of occurrence of the term t in $T(g)$. The higher the entropy the greater the variety of terms and, according to the theory, the more likely the group to be bond-based as opposed to identity-based. Since not all groups have the same number of terms and the entropy value grows with the total number of terms, a *normalized entropy* value h_g, is computed normalizing by the average value of entropy for all the groups in the system with the same number of terms:

$$h_g = \frac{H(g)}{\langle H(f) \rangle_{|T(g)|=|T(f)|}}. \tag{4}$$

Reciprocity and entropy (and a small set of other measures derived from them) are combined either linearly or, if a training set is available, with a machine learning approach. Grabowicz et al. tested both methods on a sample of Flickr groups whose type has been inferred from a manual inspection of the group page. The classification of groups with the best algorithmic setting yields very accurate results (AUC = 0.88). For the sake of illustration, we report in Fig. 2 how the likelihood of a group being social (as opposed to topical) increases as reciprocity $\hat{r}_g$ and entropy h_g increase.

In summary, the research we have reviewed in this section is an illustration of how measurable signals from online media activity can be



The Nature of Social Structures, Fig. 2 Likelihood of a Flickr group to be bond-based (social), as opposed to identity-based (topical) as group reciprocity $\hat{r}_g$ and group discussions entropy h_g increase. Each point is calculated on a 50-groups statistics. As expected from the *common identity and common bond* theory, the higher the reciprocity and variety of topics discussed the higher the likelihood of a group being bond-based. (Figure adapted from Grabowicz et al. (2013))

combined under the direction of grounded sociological theories in order to quantify concepts that contribute to define the nature of social structures. In the simple case-study discussed, the description of a group's type based on the motivation of people to be part of it (topical interest vs. social bonding) surpasses the structural and semantic layers of abstraction and gives a glimpse over the pragmatics of the group dynamic.

Resource Exchange on Social Ties

There exist very little data-driven work on social ties that can provide a computational understanding that goes past their structural or semantic interpretation. On the other hand, social scientists have been working for decades to lay the foundations of general models to explain the deeper meaning of social interactions. The *social exchange* theory (Blau 1964), developed mainly by Peter Blau and Richard Emerson, is arguably the most established framework to interpret the nature of social ties. In short, the theory conceives every social dyad as a repeated set of exchanges of different types of nonmaterial *resources* transacted in an interpersonal situation, such as knowledge, social support, or manifestation of approval (Foa and Foa 1980). Only recently part of this framework has been operationalized and tested on large-scale online data by Aiello et al. (2014). In the following, we briefly summarize the approach to provide yet another example of how a computational approach grounded in social science concepts can unfold higher-level explanations of the meaning of social structures.

The method aims at discovering the types of resources exchanged in a communication network and to cluster messages by the type of resource they convey, rather than by their topical aspect. The proposed algorithm is based on a key intuition given by the social exchange theory: in a dyad, social interactions conveying a resource tend to be reciprocated with the same resource type. As an illustration, if two individuals exchange knowledge now, their next exchange will be most likely to also involve knowledge, rather than affection. This hypothesis has been previously validated for a wide range of social interactions both offline and online (Gould 2002; Antonucci et al. 1990).

More formally, the problem addressed is defined as follows.

Input a population of users U and a set of messages M where each message $m^t_{u,v} \in M$ is a textual communication between source $u \in U$ and destination $v \in U$ at time t.

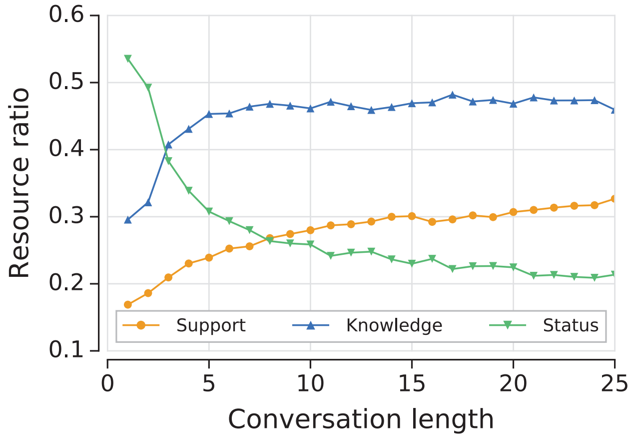
Output A probabilistic clustering of messages in M with probability of a message m to be assigned to cluster D being $p(m, D) \geq 0$.

The novel aspect of the method compared to topic models is the nature of the clusters in output. The message grouping is not done based on their topical aspects, but according to the type of social exchange those messages convey. The method is executed in four phases. First, the bag of words of each message is parsed with standard filters commonly used in information retrieval such as stopword removal, stemming, and n-gram expansion. Second, messages are probabilistically assigned to topics using matrix decomposition techniques. Specifically, Non-negative Matrix Factorization (NMF) is used, but other approaches like Singular Value Decomposition (SVD) or Principal Component Analysis (PCA) might be used (Arora et al. 2012). To select the number of topics, an iterative approach is used to select the number that minimizes the Frobenius norm of the error matrix in the decomposition. Third, a *Conversation Graph* is built; its nodes are the topics, and an edge between topics A and B is weighted by the number of times a conversation between two individuals transitions from A to B. A topical transition is simply defined as a pair of two consecutive messages sent between user u and v where message $m_1 : u \rightarrow v$ belongs to A (with highest probability) and message $m_2 : u \rightarrow v$ belongs to B.

The last step is aimed at extracting the pragmatics layer from the graph of transitions between topics. The Conversation Graph shapes the transition between classes of coherent messages during social interactions. These interactions are conceived as the realizations of underlying processes of social resources exchange. Based on the theory, it is assumed that a message that conveys a certain type of resource will most likely get a reply that conveys the same resource type. For instance, it is expected that a person who receives social support for the loss of a grieving relative ("I'm sorry for your loss") to reply in kind (if at all) with another social support interaction ("Thank you for being a good friend") rather than a status-exchange interaction ("You're such a great photographer!"). Under this interpretation, highly clustered parts of the Conversation Graph aggregate topics that carry homogeneous patterns of social exchange and will have fewer edges connecting them to the rest of the graph. This scenario is consistent with the most common definition of graph *community* (Fortunato, 2010b); therefore network community detection algorithms are applied to the Conversation Graph to discover these dense areas. The assumption is that each community contains topics whose messages carry the same resource type. Each message is probabilistically assigned to a topic and, in turn, each topic is assigned to a community (i.e., to a resource type). Therefore, a message can be described as a probabilistic vector of resource types, which is the expected output.

When tested on conversational datasets from Flickr and the bibliophile community aNobii (Aiello et al. 2010), three resource types are found. Manual inspection of messages reveals that those resources accurately depict three types of exchange that have been for long studied in social science: *knowledge exchange* (knowledge about the specific platform domains, i.e., photos and books), *status exchange* (appreciation, esteem, or admiration (Cook and Emerson 1978)), and *social support* (instrumental aid, emotional caring, or concern (House et al. 1988)).

Besides providing evidence about the accuracy of the method, Aiello et al. show that the newly generated knowledge can help to better interpret social network analysis results. One example is given about assortativity. The aNobii communication network follows an assortative pattern, i.e., people who communicate to each other receive a similar number of messages overall. The aNobii communication network is assortative but, when focusing on the set of edges that convey status, it becomes disassortative, meaning that status giving follows a hierarchical pattern: People receive status from people who have less status than them. A similar discrepancy in assortativity trends has been found in previous work when decomposing a social network in two networks based on link polarity (positive vs. negative links) (Ciotti et al. 2015). Another example is tie evolution. When computing the average ratio of messages belonging to each of the three resources for



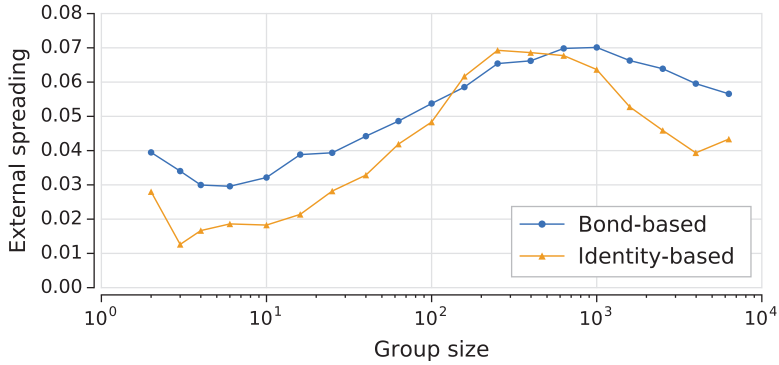
The Nature of Social Structures, Fig. 3 Average proportion of messages carrying each of the three nonmaterial resources (support, knowledge, status) for pairs of users with fixed conversation length. As a conversation thread becomes longer, the exchange of knowledge and support grows whereas status-giving messages become less frequent (Figure adapted from Aiello et al. (2014))

conversations with fixed length *n*, interesting patterns emerge (Fig. 3). Status exchange is particularly present in short conversations or, more in general, in the first stages of a conversation, after which the average tie moves to a mix of knowledge exchange and social support. It thus appears that status exchange serves to set the foundation for the future relationship, fading to the interactional background after the tie-formation stage. Such a nuanced description of the temporal evolution of a social relationship is possible only because of the ability to describe a social interaction according to their purpose, rather than their topical theme or structural position. This is why a deep computational understanding of the nature of social structures is important to advance our knowledge of societal dynamics.

Key Applications

The characterization of messages in terms of their type of social exchange opens up to a plethora of unexplored opportunities for several applications, not limited to analytics. First is user profiling: users engaged in conversations that are predominantly characterized by different resources would be presumably interested in different types of activities (e.g., socialization vs. item consumption). Second is link profiling: dyads exchanging different social resources might react differently to signals. For example, when considering a process of information diffusion (e.g., diffusion of product ads via viral marketing), considering the knowledge, status, or social support networks may yield very different results. Last, there are open opportunities for the summarization of social relationships. For example, Facebook's friendship page displays a relationship between two connected users with a timeline of their shared experiences. The form of tie decomposition in resources that we have illustrated would allow a different way of summarizing a social link, e.g., "based on their conversations, Alice and Bob's relationship is made 30% by knowledge exchange, 20% by status giving and 50% by social support."

Group characterization has important practical implications, for example, when studying information diffusion. Martin-Borregon et al. (2014) built on the operationalization of the Common Identity and Common Bond theory to verify if the span of information cascades is affected by the nature of groups these cascades are originated from. To do that, they study how Flickr photos uploaded to social and topical groups spread along social ties. Spreading is estimated by the analysis of temporal sequences of users marking a



The Nature of Social Structures, Fig. 4 Average values external spreading (i.e., proportion of infected nodes that reside outside the group) measured for information diffusion phenomena that are originated inside groups of different types (bond-based and identity-based) (Figure adapted from Martin-Borregon et al. (2014))

photo as one of their "favorite." With the goal of checking whether a photo that is uploaded in a group pool has a diffusion that is predominantly restricted to that group or spreads beyond the group boundaries, they measure the *external spreading* of a cascade simply by counting the number of *infected* users (i.e., those reached by the cascade) who reside outside the group over the total number of nodes infected by that cascade. They find that social groups ease the information spreading across the group boundaries more than topical groups, as shown in Fig. 4.

Future Directions

Research aimed at producing a more detailed characterization of the nature of social structure is still in its infancy. Here we have shown just a few examples of how this field has been developing in the last years; there are many open opportunities to push the boundaries of our knowledge in this area.

Groups might be the element to bridge micro- and macroanalysis of social systems. Social networks are complex systems, where relationship between atomic components give rise to an emergent behavior that cannot be inferred or modeled directly from the composition of the individual parts. Complex processes in networks have been studied in several fields including physics, biology, and computer science but also in social sciences, where the duality (and often the incoherence) between the behavior of an individual actor or of its interpersonal dyadic relations and the behavior of masses of people is still an important subject of investigation. In his book (Blau 1964), Peter Blau commented on this challenge:

> The problem is to derive the social processes that govern the complex structures of communities and societies from the simpler processes that pervade the daily intercourse among individuals and their interpersonal relations.

Later, in an updated introduction to the same book, he states:

> I thought that this microsociological theory could serve as a foundation for building a macrosociological theory; I no longer think this is true. The reason is that microsociological and macrosociological theories require different approaches and conceptual schemes, and their distinct perspective enrich each other.

Groups fall in between the micro- and macroscale: they are mesoscopic social structures that are born from the composition of individual drives – people trying to build their social identity or to create social bonds – but they have also a role in explaining global network phenomena. The important role of groups in bridging different scales motivates even more the need for a nuanced characterization of their multiple facets.

Turning to social ties, a more systematic way to represent their nature could lead to open whole new research fields. The representation of a social tie as a sequence of individual exchanges naturally leads one to the idea of understanding social ties as strings of interactions. With this understanding, one could use insights from theoretical computer science to establish the computational properties of social rituals. Indeed, this idea has already been leveraged by DeDeo (2013), who gives evidence of the insufficiency of finite-state machines for the description of social interactions. The ultimate goal of such analysis is the unpacking of "culture" as a formal, computational concept. If we see social ties as interactional sequences, then we may understand the resources exchanged in interpersonal communications as the "grammar of society" (Bicchieri 2006) – in other words, the bits of "source-code" that prescribe how individuals are to act in a certain situation. With another analogy, we can imagine the emergence of a line of research we could name *social chemistry*, where fundamental elements of human interactions are combined in higher-level molecules which in turn build the societal dynamics we observe everyday.

Acknowledgments We thank Bogdan State, Rossano Schifanella, Przemyslaw Grabowicz, David Martin-Borregón, Victor Eguiluz, Alejandro Jaimes, and Ricardo Baeza-Yates for the research work they have conducted jointly with the author and parts of which are summarized in this entry.

Cross-References

- ▶ Community Evolution
- ▶ Group Representation and Profiling
- ▶ Models for Community Dynamics
- ▶ Network Science
- ▶ Node Centrality
- ▶ Online Communities
- ▶ Opinion Diffusion and Analysis on Social Networks
- ▶ Semantic Social Networks
- ▶ Semantic Social Networks Analysis
- ▶ Sentiment Analysis in Social Media
- ▶ Social Groups in Crowd
- ▶ Topic Modeling in Online Social Media, User Features, and Social Networks for
- ▶ Tracking Dynamic Community Evolution in Social Networks

References

Aiello LM, Barrat A, Cattuto C, Ruffo G, Schifanella R (2010) Link creation and profile alignment in the aNobii social network. In: SocialCom: proceedings of the second IEEE international conference on social computing, Minneapolis, MN, USA

Aiello LM, Barrat A, Schifanella R, Cattuto C, Markines B, Menczer F (2012) Friendship prediction and homophily in social media. ACM Trans Web 6(2):1

Aiello LM, Schifanella R, State B (2014) Reading the source code of social ties. In: WebSci: proceedings of the 2014 ACM conference on web science. ACM, Bloomington, IN, USA

Antonucci T, Fuhrer R, Jackson J (1990) Social support and reciprocity: a cross-ethnic and cross-national perspective. J Soc Pers Relat 7(4):519

Arora S, Ge R, Moitra A (2012) Learning topic models–going beyond svd. In: FOCS: IEEE 53rd annual symposium on foundations of computer science, New Brunswick, NJ, USA

Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. In: KDD: proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, Philadelphia, PA, USA

Backstrom L, Kumar R, Marlow C, Novak J, Tomkins A (2008) Preferential behavior in online groups. In: WSDM: proceedings of the international conference on Web search and data mining. ACM, Palo Alto, CA, USA

Backstrom L, Kleinberg J, Lee L, Danescu-Niculescu-Mizil C (2013) Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In: WSDM: proceedings of the international conference on Web search and data mining. ACM, Rome, Italy

Baldassarri A, Barrat A, Capocci A, Halpin H, Lehner U, Ramasco J, Robu V, Taraborelli D (2008) The berners-lee hypothesis: power laws and group structure in flickr. In: Alani H, Staab S, Stumme G (eds) Social Web communities, Dagstuhl seminar proceedings. Schloss Dagstuhl, Dagstuhl, Germany

Barrat M, Barthelemy R, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proc Natl Acad Sci U S A 101(11):3747

Bicchieri C (2006) The grammar of society: the nature and dynamics of social norms. Cambridge University Press, Cambridge

Boyd D, Golder S, Lotan G (2010) Tweet, tweet, retweet: conversational aspects of retweeting on twitter. In: HICSS: 43rd Hawaii international conference on system sciences. IEEE, Honolulu, HI, USA

Bramsen P, Escobar-Molano M, Patel A, Alonso R (2011) Extracting social power relationships from natural Language. In: Human Language technologies, vol 1. Association for Computational Linguistics

Butler B (1999) When a group is not a group: an empirical examination of metaphors for online social structure. PhD thesis, Carnegie Mellon University

Celli F, Rossi L (2012) The role of emotional stability in twitter conversations. In: Workshop on semantic analysis in social media. Association for Computational Linguistics, Avignon, France

Ciotti V, Bianconi G, Capocci A, Colaiori F, Panzarasa P (2015) Degree correlations in signed social networks. Physica A 422:25

Cook KS, Emerson RM (1978) Power, equity and commitment in exchange networks. Am Sociol Rev 43(5):721

Correa T, Hinsley AW, de Zúñiga HG (2010) Who interacts on the Web? The intersection of users' personality and social media use. Comput Hum Behav 26(2):247

Cox A, Clough P, Siersdorfer S (2011) Developing metrics to characterize Flickr groups. J Am Soc Inf Sci Technol 62(January):493

Danescu-Niculescu-Mizil C, Lee L, Pang B, Kleinberg J (2012) Echoes of power: Language effects and power differences in social interaction. In: WWW: proceedings of the 21st international conference on World Wide Web. ACM, Lyon, France

De Choudhury M (2009) Modeling and predicting group activity over time in online social media. In: HT: proceedings of the 20th ACM conference on hypertext and hypermedia. ACM, Torino, Italy

De Choudhury M, Sundaram H, John A, Seligmann DD (2009) What makes conversations interesting?: themes, participants and consequences of conversations in online social media. In: WWW: proceedings of the 18th international conference on World Wide Web. ACM, Torino, Italy

DeDeo S (2013) Collective phenomena and non-finite state computation in a human social system. PLoS One 8(10):e75818

Dunbar RIM (1998) The social brain hypothesis. Evol Anthropol 6:178

Foa EB, Foa U (1980) Resource theory: interpersonal behavior as exchange. In: Social exchange: Advances in theory and research. Plenum Press, New York

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75

Gautreau A, Barrat A, Barthélemy M (2009) Microdynamics in stationary complex networks. Proc Natl Acad Sci 106(22):8847

Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: CHI: proceedings of the SIGCHI conference on human factors in computing systems. ACM, Boston, MA, USA

Gloor PA, Zhao Y (2006) Analyzing actors and their discussion topics by semantic social network analysis. In: IV: proceedings of the conference on information visualization. IEEE, Baltimore, MD, USA

Goncalves B, Perra N, Vespignani A (2011) Modeling users' activity on twitter networks: validation of Dunbar's number. PLoS One 6(8):e22656

Gonçalves P, Araújo M, Benevenuto F, Cha M (2013) Comparing and combining sentiment analysis methods. In: COSN: proceedings of the first ACM conference on online social networks. ACM, Boston, MA, USA

Gould RV (2002) The origins of status hierarchies: a formal theory and empirical test. Am J Sociol 107(5):1143

Grabowicz PA, Eguíluz VM (2012) Heterogeneity shapes groups growth in social online communities. Europhys Lett 97(2):28002

Grabowicz PA, Ramasco JJ, Moro E, Pujol JM, Eguiluz VM (2012) Social features of online networks: the strength of intermediary ties in online social media. PLoS One 7(1):e29358

Grabowicz PA, Aiello LM, Eguiluz VM, Jaimes A (2013) Distinguishing topical and social groups based on common identity and bond theory. In: WSDM: proceedings of the sixth ACM international conference on Web search and data mining. ACM, Rome, Italy

Granovetter MS (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380

Harper M, Frankowski D, Drenner S, Ren Y, Kiesler S, Terveen L, Kraut R, Riedl J (2007) Talk amongst yourselves: inviting users to participate in online conversations. In: IUI: proceedings of the 12th international conference on intelligent user interfaces. ACM, Honolulu, HI, USA

Holme P, Saramäki J (2012) Temporal networks. Phys Rep 519(3):97

Holmes P, Cox AM (2011) Every group carries the flavour of the admins. Leadership on flickr. Int J Web Based Commun 7(3):376

Honeycutt C, Herring S (2009) Beyond microblogging: conversation and collaboration via Twitter. In: HICSS: 42nd Hawaii international conference on system sciences. IEEE Computer Society, Honolulu, HI, USA

House J, Umberson D, Landis K (1988) Structures and processes of social support. Annu Rev Sociol 14:293

Java A, Song X, Finin T, Tseng B (2007) Why we twitter: Understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM, San Jose, CA, USA

Kairam SR, Wang DJ, Leskovec J (2012) The life and death of online groups: predicting group growth and longevity. In: WSDM: proceedings of the fifth ACM international conference on Web search and data mining. ACM, Seattle, WA, USA

Kim S, Bak J, Oh A (2012) Do You feel what I feel? Social aspects of emotions in Twitter conversations. In: ICWSM: proceedings of the 6th international AAAI conference on Weblogs and social media, Dublin, Ireland

Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks. J Complex Netw 2(3):203

Kooti L, Aiello M, Grbovic M, Lerman K, Mantrach A (2015) Evolution of conversations in the age of email overload. In: WWW: proceedings of the 24th international conference on World Wide Web. ACM, Florence, Italy

Kumar R, Mahdian M, McGlohon M (2010) Dynamics of conversations. In: KDD: proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, Washington, DC, USA

Kunegis J, Lommatzsch A, Bauckhage C (2009) The slashdot zoo: mining a social network with negative edges. In: WWW: proceedings of the 18th international conference on World Wide Web. ACM, Madrid, Spain

Kunegis J, Preusse J, Schwagereit F (2013) What is the added value of negative links in online social networks? In: WWW: proceedings of the 22nd international conference on World Wide Web. ACM, Rio De Janeiro, Brazil

Laine M, Ercal G, Luo B (2011) User groups in social networks: an experimental study on youtube. In: HICSS: 44th Hawaii international conference on system sciences. IEEE Computer Society, Honolulu, HI, USA

Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. Phys Rev E 78:046110

Leskovec J, Huttenlocher D, Kleinberg J (2010) Predicting positive and negative links in online social networks. In: WWW: proceedings of the 19th international conference on World Wide Web. ACM, Raleigh, NC, USA

Ludford PJ, Cosley D, Frankowski D, Terveen L (2004) Think different: increasing online community participation using uniqueness and group dissimilarity. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, Vienna, Austria

Martin-Borregon D, Aiello LM, Grabowicz P, Jaimes A, Baeza-Yates R (2014) Characterization of online groups along space, time, and social dimensions. EPJ Data Sci 3(1):1

Miller D, Edwards WK (2007) Give and take: a study of consumer photo-sharing culture and practice. In: CHI: proceedings of the SIGCHI conference on Human factors in computing systems. ACM, San Jose, CA, USA

Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: IMC: proceedings of the 7th ACM SIGCOMM conference on internet measurement. ACM, San Diego, CA, USA

Negi S, Chaudhury S (2012) Finding subgroups in a flickr group. In: ICME: proceedings of the 2012 I.E. international conference on multimedia and expo. IEEE, Melbourne, Australia

Negoescu RA, Gatica-Perez D (2008a) Analyzing flickr groups. In: CIVR: proceedings of the international conference on content-based image and video retrieval. ACM, Niagara Falls, Canada

Negoescu RA, Gatica-Perez D (2008b) Topickr: flickr groups and users reloaded. In: MM: proceedings of the 16th ACM international conference on multimedia. ACM, New York, Vancouver, Canada

Negoescu RA, Gatica-Perez D (2010) Modeling Flickr communities through probabilistic topic-based analysis. Trans Multimedia 12(5):399

Negoescu R-A, Adams B, Phung D, Venkatesh S, Gatica-Perez D (2009) Flickr hypergroups. In: MM: proceedings of the 17th ACM international conference on multimedia. ACM, Beijing, China

Newman ME (2003) The structure and function of complex networks. SIAM Rev 45(2):167

Nov O, Naaman M, Ye C (2010) Analysis of participation in an online photo-sharing community: a multidimensional perspective. J Am Soc Inf Sci Technol 61(3):555

Pissard N, Prieur C (2007) Thematic vs. social networks in web 2.0 communities: a case study on flickr groups. In: Algotel conference, Ile d'Oléron, France

Plutchik R (1980) Emotion: a psychoevolutionary synthesis. Harper and Row, New York

Porter E (2004) A typology of virtual communities: a multi-disciplinary foundation for future research. J Comput-Mediat Commun 10(1):00–00

Prentice A, Miller DT, Lightdale JR (1994) Asymmetries in attachments to groups and to their members: distinguishing between common-identity and common-bond groups. Personal Soc Psychol Bull 20(5):484

Prieur C, Cardon D, Beuscart J-S, Pissard N, Pons P (2008a) The strength of weak cooperation: a case study on flickr. CoRR. abs/0802.2317

Prieur C, Pissard N, Beuscart J, Cardon D (2008b) Thematic and social indicators for Flickr groups. In: ICWSM: proceedings of the 2nd international AAAI conference on Weblogs and social media, Seattle, WA, USA

Purohit Y, Ruan D Fuhry S (2014) Parthasarathy, and A. Sheth. On understanding divergence of online social group discussion. In: ICWSM: proceedings of the 8th international AAAI conference on Weblogs and social media, Ann Arbor, MI, USA

Spertus E, Sahami M, Buyukkokten O (2005) Evaluating similarity measures: a large-scale study in the orkut social network. In: KDD: proceedings of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining. ACM, Chicago, IL, USA

Tajfel H (1981) Human groups and social categories. Cambridge University Press, Cambridge

Tang L, Wang X, Liu H (2011) Group profiling for understanding social structures. ACM Trans Intell Syst Technol 3(1):1

Tchokni S, Séaghdha DÓ, Quercia D (2014) Emoticons and phrases: status symbols in social media. In: ICWSM: proceedings of the 8th international AAAI conference on Weblogs and social media, Ann Arbor, MI, USA

Van House NA (2007) Flickr and public image-sharing: distant closeness and photo exhibition. In: CHI: extended abstracts on human factors in computing systems. ACM, San Jose, CA, USA

Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: WOSN: proceedings of the 2nd ACM workshop on online social networks. ACM Press, Barcelona, Spain

Wang J, Zhao Z, Zhou J, Wang H, Cui B, Qi G (2012) Recommending flickr groups with social topic model. Inf Retr 15(3–4):278

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, New York

Welser HT, Gleave E, Fisher D, Smith M (2007) Visualizing the signatures of social roles in online discussion groups. J Soc Struct 8(2):564

Wilson C, Boe B, Sala A, Puttaswamy KP, Zhao BY (2009) User interactions in social networks and their implications. In: EuroSys: proceedings of the fourth ACM european conference on computer systems. ACM, Nuremberg, Germany

Xiang R, Neville J, Rogati M (2010) Modeling relationship strength in online social networks. In: WWW: proceedings of the 19th international conference on World Wide Web. ACM, Raleigh, NC, USA

Yang J, Leskovec J (2015) Defining and evaluating network communities based on ground-truth. Knowl Inf Syst 42(1):181–213

Recommended Reading

Austin J (1975) How to do things with words. Oxford university press, Oxford

Blau P (1964) Exchange and power in social life. Transaction Publishers, New Brunswick

Emerson R (1962) Power-dependence relations. Am Sociol Rev 27:31

Emerson R (1976) Social exchange theory. Annu Rev Sociol 2:335