Understanding Community Dynamics in Online Social Networks

Hari Sundaram[‡]   **Yu-Ru Lin**[†]
Munmun De Choudhury[§]   **Aisling Kelliher**[‡]
[‡]Arizona State University, Tempe, AZ 85281
[†]Northeastern University, Boston, MA 02115
[§]Rutgers, The State University of New Jersey, NJ 08901
[§]m.dechoudhury@rutgers.edu, [†]yuruliny@gmail.com
[‡]{hari.sundaram, aisling.kelliher}@asu.edu

Abstract

Social network systems are significant scaffolds for political, economic and socio-cultural change. This is in part due to the widespread availability of sophisticated network technologies and the concurrent emergence of rich media websites. Social network sites provide new opportunities for social-technological research. Since we can inexpensively collect electronic records—over extended periods—of social data, spanning diverse populations, it is now possible to study social processes on a scale of tens of million individuals. To understand the large-scale dynamics of interpersonal interaction and its outcome, this article links the perspectives in the humanities for analysis of social networks to recent developments in data intensive computational approaches. With special emphasis on social communities mediated by network technologies, we review the historical research arc of community analysis, as well as methods applicable to community discovery in social media.

1   Introduction

Today, social networks are significant catalysts of political, social and cultural change. The catalysis is in part due to the availability of sophisticated Internet based communication technologies (collectively known as Web 2.0), and due to the emergence of rich media websites (Facebook, YouTube, Flickr are well known examples). These websites allow a distributed network of participants to communicate via public comments or private messages, as well as share rich media content, including images and videos. Social networks evolve around communication on shared

1

content. The conversations catalyzes social processes: diffusion of ideas, cultural bias, and community evolution. In the political realm, for example, social networks have been widely used as a tool to organize — the 2008 elections in the United States, and the recent developments in the Middle-East in spring 2011, are examples. These networks have altered our notions of social interaction, including friendship, and how we interact with strangers. Finally, the networks have a strong cultural influence — for a significant number of young people, these networks have become the primary source of news, and entertainment.

Social network sites provide new research opportunities for social-technological and scientific communities. Instead of focusing on longitudinal studies of relatively small groups — participant observation [1] and surveys— today, researchers can study social processes including information diffusion or community emergence at *very large scales*. The study at large scales is made possible by collection of electronic records of social data, spanning diverse populations, over extended time periods. Importantly, we can do so at a comparatively low cost, requiring little human supervision. The result: study of social processes on a scale of tens of millions individuals, impossible just a few years ago. An analysis of conversations within social networks, for example, provides insights into human behavior at multiple levels, including temporal and topological levels. In particular, it helps researchers understand large scale online communities as an emergent property of social interaction.

Community discovery in a social network has many applications. These include expertise finding and neighborhood query, and behavioral prediction. The structure of a community, which accounts for inherent dependencies between individuals in a social network, can help us understand the behavioral dynamics of individuals. Through characterizing multidimensional interpersonal relationships and an individual's interests, community analysis can provide a quantitative summary of key factors related to word-of-mouth communications: tie strength, homophile, and source credibility. We can use community analysis to organize and to track content in online social media. There are significant opportunities for businesses: in addition to understanding user behavior and comments for better product design, businesses can take advantage of sentiment analysis of comments to proactively address negative commentary. In an enterprise setting, we can predict users future interests — in particular, documents — through community structure extracted from multiple interpersonal relationships, including formal collaboration, informal

communication and sharing [2].

We review and connect, in this article, the methodologies developed in multiple disciplines, including the humanities, and network science, to recent developments in data intensive computational approaches. In particular, we shall examine formation of online communities.

Given limited space for this article, a comprehensive survey of community detection methods would be dense at best, and an incomplete description at worst, of the problem area. For a comprehensive description, we point the interested reader to an excellent recent review on community detection [3].

Our focus—in this paper—is to explain an under-appreciated aspect of the problem area: linkages amongst multiple disciplinary perspectives on community formation and detection. We specifically link perspectives from sociology, computer mediated communication, and network science, to data intensive computational perspectives. We show how contemporary methods based on clustering can be adapted to include temporal and contextual aspects emphasized by other disciplines. This linkage between the different disciplines nicely complements review papers, where the focus is on careful examination of different quantitative methods.

In the next section, we shall discuss in detail the historical research arc of community analysis. We plan to discuss computational methods for community discovery in Section 3 on page 7. In Section 4 on page 14, we shall present some example applications for community discovery. Finally in Section 5 on page 16, we present our conclusions.

2   What is a Community?

In this section we discuss the formation of an important macroscopic structure — a community — though interaction amongst individuals. We first discuss the definition of a community, including definitions that geographically bound the notion of a community. Then in Section 2.1 on page 5, we discuss virtual communities, and how they are distinct from chance encounters between people. Finally in Section 2.2 on page 6, we present a powerful network-based representation of social interaction.

The concept of a 'community' affords many definitions [4]. Our understanding of a 'community' is informed by critical research in several fields including anthropology, sociology, political

science and the wider humanities. A traditional understanding of community is strongly aligned with the notion of a neighborhood or a village, where interpersonal ties are considered to be locally bounded [5]. Consequently, concerns about loss of community have been raised when observers cannot find much solidary local behavior and sentiments. This framing of a community is challenged by contemporary scholars who seek to study interpersonal relations in the form of networks that are both local and geographically unbound. Historian Benedict Anderson critiques the constraint of community analysis to localized, face-to-face interaction in his description of the nation state, which he defines as an imagined community [6]. According to Anderson, a national imagined community is a socially constructed mental image where members "will never know most of their fellow-members, meet them, or even hear of them, yet in the minds of each lives the image of their communion." Anderson's definition of nations relies on an extended conception of communities, which he expressed in several ways: shared consciousness, technology condition (e.g. print), and technology enhancement (e.g. census, maps and museums).

Sociologist Barry Wellman extends the notion of community to encompass more general networks of interpersonal ties that provide sociability, support, information, a sense of belonging and social identity [7]. This framing of the concept of 'community,' addresses concerns about loss of community in the absence of significant solidary behavior and sentiments solely within a locality [5]. This broader view is shared by other disciplines, including anthropology, where recent studies have suggested that "a more fluid concept of community fits well within ethnographic explorations in multi-sited situations with complex, spatially diverse communities". In these studies, communities are observed to be fluid and to be flexible, and may be based on a wide range of cultural interests and social affiliations.

The characteristics of community have also been examined within the field of situated cognition. According to Dewey, an individual's actions will always be interrelated to all others within certain social medium that helps form the individual's membership in a community. Once membership is established, the individual begins to share the knowledge possessed by the group. This shared experience forms an emotional tendency: it motivates individual behavior in a way to create purposeful activity, thus evoking meaningful outcomes. These studies suggest that the behavioral dynamics of individuals occur under complex, social conditions that simultaneously give rise to the community structure (i.e. the "dense cluster" or "community membership"). While the

conditions may be ambiguous, situated cognition theorists have suggested that "artifacts [hold] historic and negotiated significance within a particular context". Lemke [8] described community ecology as follows: "they have a relevant history, a trajectory of development in which each stage sets up conditions without which the next stage could not occur," and "the course of their development depends in part on information laid down (or actively available) in their environments from prior (or contemporary) systems of their own kind".

In the next section — moving away from the geographically bound groups — we discuss virtual communities, including conditions for online community formation.

2.1 Virtual Communities

In Computer-Mediated Communications (CMC) research, investigators have shifted attention away from officially-defined group or geographical boundaries toward conditions or characteristics for online community formation. Preece [9] provides a working definition of online community comprising the following elements: "people who interact for their own needs or perform special roles; a shared purpose such as an interest, need, information exchange, or service that provides a reason for the community; policies that guide people's interactions; computer systems which support and mediate social interaction and facilitate a sense of togetherness." This definition seeks to provide a framework to guide developers in making operational decisions for designing and building online communities.

Garfinkel's observation on the necessity of mutually observable actions within community members [10] has influenced views in CMC research, on how "interactivity" forms a social reality. According to Dourish [11], interaction involves presence (some way of making the actors present in the locale) and awareness (some way of being aware of the others presence). An action community, according to Dourish, is one where members share common understandings through reciprocal actions.

A virtual community has several characteristics that distinguish it from a chance meeting of people. Jones [12] conceptualized the notion of a virtual community based on the definition of a virtual settlement (the place, or cyberplace, where a virtual community forms). He identified four necessary characteristics of a virtual community: interactivity, communicators, a publicly shared

5

mediated communication place and sustained membership. The interactive nature of virtual communities distinguishes them from a group. A virtual community is distinguished by long term, meaningful conversations among members. McMillans socio-psychological model [13] hypothesizes the presence of four dimensions for a sense of community to emerge: feelings of membership, feelings of influence, integration and fulfillment of needs, and shared emotional connections. Blanchard [14] extends the work of Jones [12] to analyze the notion of virtual community among weblogs, based on McMillans model. Based on a survey of blog readers, Blanchard argued that a sense of community is an essential characteristic that distinguishes a virtual community from a mere virtual group.

There is considerable debate about the authenticity and value of online, virtual or computer-mediated communities. For some, contemporary technological advances are resulting in the loss of many "third places" — places for socializing outside of work and the home, where community members gather with a sense of belonging and engage in easy conversation with friends and acquaintances. These same advances are considered very differently by others, who instead see online communities as offering an alternative and vibrant third place for interpersonal communication and social support.

In the next section, we discuss network analysis, a powerful perspective on how to represent social interaction. While the network analysis perspective predates contemporary technological advances, current computational data-mining techniques for large scale analysis of data obtained from social network websites are built upon ideas from classical network analysis.

2.2   Network analysis

The small world phenomenon first identified by Stanley Milgram can be understood as a harbinger for the field of social network analysis [15]. Wellman [5] formally proposed social network analysis as a way to study community without a locally-confined presumption or other a priori analytic constraints. Social network analysis starts with a set of network members (called nodes) and a set of ties that connects some or all nodes [16]. "The utility of the network approach is that it does not take as its starting point putative neighborhood solidarities nor does it seek primarily to find and explain the persistence (or absence) of solidary sentiments. Thus the network approach at-

tempts to avoid individual-level research perspectives, with their inherently social psychological explanatory bases that see internalized attitudes as determining community relations" [5].

Social network analysis has been popularized by Granovetter's "strength of weak ties" thesis [1]. He showed that that job-seekers in Boston found their "weak" connections to be more useful in the job market than connections signifying "strong" bonds of close friendship and kinship. The study has motivated considerable research interest into the role of 'ties': from analysis of predefined social boundaries to study of interpersonal relationships. In Granovetter's study, tie strength was influenced the following: the amount of time spent, emotional intensity, intimacy and reciprocity of services. In other work, tie strength has been considered as "a multidimensional construct that represents the strength of the dyadic interpersonal relationship in the context of social networks"; the multiplex ties offer diversified support to people in a community [17]. Granovetter's "social embeddedness" theory [18] suggests that the choices available to a person, depends on their integration within dense clusters or multiplex relations of social networks. Social embeddedness in cohesive structures, for example, can lead people to make similar political contributions. Social network analysis has now been used in a variety of research areas, including the spread of diseases and information, the sociology of organizations, and Internet studies.

In this section, we presented several different views of the concept of a community. We showed how our understanding evolved from one that limited communities geographically to a virtual community built upon notions of a virtual settlement. We concluded by discussing a powerful representation of a social interaction: a network. Next, we discuss computational methods, which are based on a network representation of social interaction, to extract communities.

3 Community formation and evolution

Identification of communities as cohesive subgroups of individuals within a network, where cohesive subgroups are defined as "subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties" [16], is an important research topic in social network analysis. This is because social network analysis does not presume a prior solidary local bounds that organize peoples interpersonal relationship. Newman [19] gives a broad review of important findings and concepts in network research, including degree-distribution, small-world effect and commu-

nity structure.

In the next section, we present an operational definition of a community, guided by multi-disciplinary scholarship from the humanities, computer mediated communication and network science. Then, in Section 3.2 on the next page, we discuss modes of interaction in online social networks. These modalities are responsible for people to become gradually aware of each other — laying the foundation for the emergence of a community. Finally, in Section 3.3 on page 11, we discuss methods for community detection, including clustering methods and extensions to clustering that incorporate temporal and contextual information.

3.1 An Operational Definition

To enable analysis grounded on the predominant existing social and network methodologies, we provide an operational definition of a community is as follows: *a community refers to a cluster of people interacting with one another in a coherent manner.* The interactions can be explicit (e.g. direct email exchange between two users), or implicit (e.g. two users bookmark the same document). Garfinkel's notion of observable interactions [10], which are central to our definition, have two important characteristics: temporal and contextual coherence. Interactions are said to be temporally coherent if the degree of interaction in the interacting pair is sustained over a period of time. Two people exchanging emails, for example, over a sustained period of time count as being temporally coherent. Interactions are contextually coherent if they have similar interaction context: time, location, people or objects associated with the interaction. When people becoming gradually aware of each other, through observations of coherent interactions, a community begins to emerge. People who bookmark articles, for example, that share the same context would be noted as being contextually coherent. High school students from a particular school bookmarking websites related to a common class project, is a specific example.

The social significance of the communities relies on several assumptions. First, community members develop awareness of one another through observing coherent interactions. The observable temporally and contextually coherent interactions among people, give rise to other non-observable interpersonal properties, including shared consciousness and emotional bonds. Second, we assume there exists a two-way communication mechanism that allows coherent interac-

tions to take place and to be observed.

3.2   Interaction Modes in Social Media

Social interaction is central to community formation and to the evolution of social systems. It is the process by which participating individuals create and share information with one another in order to reach a mutual understanding. Over the years, numerous empirical studies [20] on online social communication processes have revealed that properties of the associated social system, including the network structure and dynamics, can determine the outcome of important social and economic relationships.

Social media, which enable mutual observability and two-way communications, are Internet-based tools that enable people to communicate and interact with each other in various media forms including text, images, audio and video. Social media sites offer many different ways in which end users can interact with the system. Various interactions frameworks allow users to asynchronously communicate with friends across the globe, by sharing media objects and posting commentary and web links. We now review the different forms of communication amongst users.

1. *Messages.* Social websites, including MySpace and Facebook, allow users to post short messages on their friends' profiles. These messages are typically short and publicly viewable to the set of friends common to both users.

2. *Blog comments / replies.* Blogging websites, including Engadget, Huffington Post, Slashdot, Mashable and MetaFilter, afford users the ability to comment and to reply to their friends' posts. An analysis of communication in these blogs provides substantial evidence of back and forth communication among sets of users.

3. *Conversations around a shared media object.* Many social websites allow users to share media objects with their local network. On Flickr, for example, a user can upload a photo viewable to her contacts via a feed. YouTube allows users to upload videos corresponding to different topics. Both social media sites support rich communication activity — via comments — around the uploaded media object. An analysis of the comments reveals a conversational structure, indicative of dialogue among users.

4. *Micro-blogging.* We define a communication modality based on user micro-blogging. Micro-blogs, popularized by Twitter are very short posts. The micro-blog post — known as a "tweet" on Twitter — often takes conversational form. This is because tweets contain syntax to direct posts at specific users. Moreover, Twitter supports the "RT" or re-tweet feature, which enables users to repost tweets that they receive. Thus, information can propagate from one personal network to another. Hence micro-blogging activity can be considered as an active medium for interaction.

5. *Social actions.* Social media sites support indirect forms of awareness. Certain social media sites, including Digg and del.icio.us, offer a different communication modality — users participate in a variety of social actions. Digg, for example, allows users to vote on (or rate) shared articles, typically news, via an action called "digging". The "like" feature on Facebook is another example where users can "like" other user statuses, photos, videos and shared links. Such social action often acts as a proxy for communication activity. This is because first, the social action is publicly observable, and second it supports the formation of bonds amongst users.

6. *Check-in services.* Recently, location-based online social networking applications have emerged, where users share their current location instantly by checking-in on websites such as Foursquare, Facebook, etc. Location-based social networks adds an important dimension to online interactions.

While social media is creates sufficient context for the formation of communities, it is not necessary — Anderson's depiction of the nation state an imagined community, is an example.

In network analysis, graphs are a natural way to represent two-way interaction amongst users in a social network. In these graphs, each node represents a user, and an edge can represent communication or more generally interaction, between a pair of users. As a specific example, an edge between two users can be indicative of communication and where the edge weight is proportional to the number of messages exchanged. In a graph based representation, all edges have the same meaning. In real-world social networks, two people may be connected via multiple relations. When multiple relations with multiple meanings exist within a network, we can use multi-graphs as a representation — where multiple edges can exist between any two people.

3.3 Methods for Community Detection

Community detection algorithms identify the modular structure of a network, where nodes represent individuals and where links represent the interaction or similarity between individuals. Intuitively, modules or communities are subset of nodes within which the links are dense and between which the links are sparse [21]. Many graph-based approaches, including those based on analysis of cliques, degree, and matrix-perturbation, have been proposed to extract cohesive subgroups from social networks [16]. Examples of detected communities range from communities of scientists working on similar areas of research [22] to authors of home pages who have some common interests. See Fortunato [3] for a comprehensive review.

We now discuss clustering methods to extract communities, followed by extensions — social-context, temporal and relational — to clustering techniques.

3.3.1 Clustering and community discovery

The algorithms for community identification are closely related to the family of algorithms for clustering. The goal of clustering is to discover groups of similar objects within the data. Each cluster (i.e. group), consists of objects that are similar to one another within the same cluster, and dissimilar to the objects in other clusters.

There are two key aspects to the mathematical formulation of any clustering technique: a measure of similarity (distance function) and an objective function (clustering criteria). The distance function and the objective function are chosen based on the grouping purpose, including to discover any underlying structure or to summarize features of the data. Methods for clustering (see [23] for a brief review) include: hierarchical clustering, partitioning, graph clustering methods , modularity based approachand block models. In hierarchical clustering, the method recursively finds nested clusters in either agglomerative (bottom-up) or divisive (top-down) way, e.g. single-link and complete-link methods. The goal of partition methods is to partition the data into a fixed number of clusters — K-means or via Expectation-Maximization.

Community identification can be considered to be clustering, in the sense that it involves a distance function and a clustering objective function, and generates a clustering assignment for each person and object to a set of clusters. While there are similarities between community extrac-

11

tion and clustering analysis, community extraction focuses on the pairwise relationship between network nodes, and more generally, the network topology.

Research on community discovery includes measures for quantifying community structure (including the clustering coefficient [19]) and techniques for community extraction. A variety of methods for extracting community structure have been proposed including *modularity* based methods [21], flow or *graph cut* based methods [24], spectral clustering or graph Laplacian based methods, and information-theoretic models [25]. Community extraction techniques have been used to study dynamic properties of communities in empirical networks [26].

One of the main challenges with a clustering framework is in cross-validating the resulting clusters. While there are many methods proposed for validating the resulting clusters, including conductance, average clustering coefficient [19], the absence of ground truth datasets complicates validation. It is entirely possible to obtain clusters that satisfy cluster validation criteria, but some of the clusters may be false communities — unmoored in real human interactions. In the next section we discuss extensions to clustering by incorporating characteristics of real-world social interaction.

3.3.2 Extensions to Clustering: Incorporating Social Context

Clustering-based methods for community detection need to account for interactions with the following characteristics: social context, temporal coherence, and contextual coherence. As mentioned in Section 2.1 on page 5, these characteristics are consistent with Garfinkel's observation on the necessity of mutual awareness, and Jones' work on the virtual community.

We formulate community discovery as clustering — involving a distance function and a clustering objective function — to generate a clustering assignment for users and media objects. Importantly, the distance function and the clustering assignment function are designed to incorporate social [27], temporaland relational [2] constraints.

We can incorporate social context with two concepts: mutual awareness and transitive awareness. Mutual awareness refers to a relationship developed through observable interactions between two people. We can define mutual awareness computationally by contextual use of (mutually observable) links in social media (e.g. blogs). If John, for example, comments on Ana's blog

post, Ana is aware of John, but John cannot be certain that Ana is aware of him, if his comment is unread. Subsequently, if Ana comments on John's blog post, there is mutual awareness between the two. Mutual awareness can be asymmetric — the asymmetry can arise, for example, when one person is a celebrity, or is touch with more people than the other. In addition, mutual awareness strength can change over time. Transitive awareness refers to a relationship — computed via a mutual awareness measure — between two connected people on a network. We can compute transitive awareness between a connected pair of users on a social network graph, through mutual-awareness expansion. We can use a random walk based distance, with an efficient method for mutual awareness expansion, to extract communities [27].

Real-world communities are based on coherent and sustained (i.e. temporally coherent) interactions. We can develop a unified framework [28], where the community structure at a given timestep is determined both by the observed networked data and by a suitable structure prior obtained from analysis of past network data. The framework extends traditional clustering by incorporating a temporal smoothness objective function into the clustering criterion to extract communities with sustained membership. We can track community evolution from the clustering results.

In real-world social interactions, people can share an interaction context. We can extend traditional clustering to extract communities with coherent contexts. In the extended clustering framework we cluster different objects including users or keywords based on their participation in different types of similarity relationships. Two users, for example, can be similar if they read the same newspaper each morning, or be similar because like to go for a run each morning.

The query-sensitive community extraction [27] uses a filtering-based approach to re-weight the interaction graph with respect to a given query and then uses the re-weighted graph to extract communities. We can develop a framework [2], with a multi-relational clustering objective, to focus on the constantly changing and co-involving interaction contexts in online social media. The framework can represent heterogeneous social contexts in social media — multi-relational and multi-dimensional social data — with a novel relational hypergraph representation called metagraph. Tensors are a natural way to encode $n-$way relationship between entities. With this generalized objective function, different type of relations — user interaction, content similarity — are considered simultaneously, allowing us to capture evolution of both user interaction and

of content interests within communities.We can extract communities through an efficient multi-relational factorization algorithm on a given metagraph.

Although these community discovery methods are similar to two recently developed clustering techniques — evolutionary clustering [29] and relational learning [30] — the focus on temporal and contextual coherence captures the nuance in online social interaction. Hence, our approach focuses on interpretable statistics such as soft-clustering and cluster transitions, which provides measures for importance of individuals in relation to communities to which they belong, as well as the community-level interactions and evolution.

In this section, we discussed three issues. First, we presented an operational definition of a community: a community refers to a cluster of people interacting with one another in a coherent manner. Then, we discussed interaction modes in social media, which enable mutual observability and two-way communications. Finally, we discussed clustering based methods for community detection, including extensions, that incorporated temporal and contextual information. Next, we discuss applications of community detection.

4   Applications

The analysis of dynamic relationship among people, concepts and contexts within a community has several applications, including information search, expert finding, content organization and behavioral prediction.

Context-sensitive information search and recommendation. The community structure extracted from multi-relational data can be used to provide context-sensitive recommendations along any attribute. When a user is looking at a particular photo, for example, we use the relational structure to find objects likely co-occurring with the photo, and then recommend other photos, tags, and related peers. The multi-relational structure provides additional context, over the $< user, photo >$ pairs used to recommend tags in automated annotation algorithms. In particular, it allows us to select peers and context, including visual features, activities, time, that are likely more related to the current user.

Expert finding and neighborhood query. A main challenge in organizational learning is to leverage the expertise of relevant peers in an timely fashion. This requires us to find a person with

relevant and valuable expertise and one who can be easily reached as "neighbors." This person can can be reached through effective communication channels including face-to-face conversation, phone calls and instant messaging. The community structure extracted by our method, from multi-relational data, including organizational structure, daily communications, and document access, can help to identify experts located in the "neighborhood" of information seekers.

Content organization, tracking and monitoring. Social media site encourage the use and sharing of multimedia content — the rate at which such content appears in these sites, creates several challenges. First, the content in a photo stream — either for a user or a community — is typically organized in temporal order, making the exploration and browsing of content cumbersome. Second, sites, including Flickr, provide frequency based aggregate statistics. The aggregate statistics include popular tags and top contributors. These aggregates do not reveal the rich temporal dynamics of community sharing and interaction; photos or posts on "Arizona Travel," for example, exhibit seasonal patterns. Additionally, these relational semantics are easily glossed over when accessing the photo stream via a single attribute including photos, users, tags or a particular time. John typically comments, for example, on Janes photos, in particular on those photos tagged with "biking" — an opaque connection. The presence of meaningful relationships between different attributes suggests new mechanisms — based on the discovered semantic relations — for content organization and presentation. We can use multirelational community discovery to extract multi-relational time-varying structure. The extracted structure will facilitate organization, tracking and monitoring of user-generated social media content.

Behavioral prediction. Studies have shown that individual behaviors, including social embeddedness [18] and influence, usually result from mechanisms that depend on their social networks. Social embeddedness framework indicates that the choices of individuals depend the mechanism of integration in dense clusters or multiplex relations of social networks. Social embeddedness in cohesive structures, for example, can lead people to make similar political contributions. Social influence refers to changes in individual characteristics that depend on the characteristics of others to whom they are tied. The opinions of individuals, for example, may be assimilated by members within the same group. Community structure, which accounts for inherent dependencies between individuals embedded in a social network, can help us understand and predict the behavioral dynamics of individuals in the community.

5 Conclusions

In this article, we presented a broad overview of of the work on communities, including research in the humanities, network science and the computing sciences. Our emphasis was to study communities formed by social interaction in online social networks. Social networks are the catalysts of significant political, economic and cultural change. The study of these websites can provide new insights into sociological processes at an unprecedented scale — we can collect electronic social data over extended periods at comparatively low cost, requiring little resource maintenance, and span diverse populations.

We reviewed the evolution of the notion of a community: from a geographically bound understanding to virtual networks. We discussed critical elements supporting community formation: interactivity, communicators, publicly shared mediated communication space and sustained membership.

We defined a community to be a cluster of people interacting with one another in a coherent manner. Social interaction is the process by which participating individuals create and share information with one another in order to reach a mutual understanding. We specifically discussed several modes of interaction, available to users in online social networks, including actions related to communication and social actions.

In our review of community detection methods, we discussed the close relationship between community discovery and clustering. We made the argument that the computational process should be able to identify communities based on interactions that include the following characteristics: relevant to community identification, temporally coherent, and contextually coherent. Finally, we discussed several interesting applications of community discovery: context sensitive information search, expertise finding, behavioral prediction, and content organization.

There are many interesting theoretical and applied questions that remain open to further study. Much of the current work on community analysis is on historical data — carefully collected from the social network over a long period of time. In such an analysis, we are unable to account for information presented to the user, about the network, on her actions, and consequently the effects of such a presentation, on the evolution of the network. Second, the role of resource costs of interacting with the network, including time costs, on the evolution of the network are poorly

understood. This is in part due to our inability to estimate these costs, and how they vary across users in the network. New applications for community discovery include collective action problems in particular, those dealing with environmental change and reducing power consumption.

References

[1] M. S. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.

[2] Yu-Ru Lin, Jimeng Sun, Paul Castro, Ravi Konuru, Hari Sundaram, and Aisling Kelliher. Metafac: community discovery via relational hypergraph factorization. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 527–536, New York, NY, USA, 2009. ACM.

[3] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[4] G. Jr. Hillery. Definitions of Community: Areas of Agreement. *Rural Sociology*, 20:111–122, 1955.

[5] B Wellman. The network community: An introduction. *Networks in the global village*, pages 1–48, 1999.

[6] B Anderson. Imagined community. *London: Verson*, 1983.

[7] Barry Wellman. Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research*, 2001.

[8] JL Lemke. Cognition, context, and learning: A social semiotic perspective. *Situated cognition: Social, semiotic, and psychological perspectives*, pages 37–56, 1997.

[9] J Preece. *Online Communities: Designing Usability and Supporting Socialbilty*. John Wiley & Sons, Inc. New York, NY, USA, 2000.

[10] H Garfinkel. *Studies in ethnomethodology*. Polity, 1984.

[11] P Dourish. *Where the Action Is:: the Foundations of Embodied Interaction*. Mit Pr, 2001.

[12] Q Jones. Virtual-communities, virtual settlements & cyber-archaeology: A theoretical outline. *Journal of Computer Mediated Communication*, 3(3):35–49, 1997.

[13] DW McMillan. Sense of community. *Journal of Community Psychology*, 24(4):315–325, 1996.

[14] AL Blanchard and ML Markus. The experienced sense of a virtual community: Characteristics and processes. *ACM SIGMIS Database*, 35(1):79, 2004.

[15] Stanley Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[16] S Wasserman and K Faust. *Social Network Analysis: methods and applications*. Cambridge University Press, 1994.

[17] J Brown, AJ Broderick, and N Lee. Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, 21(3):2, 2007.

[18] M Granovetter. Economic action and social structure: A theory of embeddedness. *American Journal of Sociology*, 91(3):481–510, 1985.

[19] M.E.J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[20] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, 311(3):590–614, 2002.

[21] MEJ Newman and M Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):26113, 2004.

[22] M Girvan and MEJ Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821, 2002.

[23] AK Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 2009.

[24] G.W. Flake, S. Lawrence, and C.L. Giles. Efficient identification of web communities. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 150–160. ACM, 2000.

[25] M. Rosvall and C.T. Bergstrom. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18):7327, 2007.

[26] J. Leskovec, K.J. Lang, A. Dasgupta, and M.W. Mahoney. Statistical properties of community structure in large social and information networks. In *Proceeding of the 17th international conference on World Wide Web*, pages 695–704. ACM, 2008.

[27] Yu-Ru Lin, Hari Sundaram, Yun Chi, Jun Tatemura, and Belle Tseng. Blog community discovery and evolution based on mutual awareness expansion. 2007.

[28] Yu-Ru Lin, Yun Chi, Shenghuo Zhu, Hari Sundaram, and Belle L. Tseng. Analyzing communities and their evolutions in dynamics networks. *Transactions on Knowledge Discovery from Data (TKDD)*, 3(2), April 2009.

[29] D Chakrabarti, R Kumar, and A Tomkins. Evolutionary clustering. pages 554–560. ACM, 2006.

[30] B Long, ZM Zhang, and PS Yu. A probabilistic framework for relational clustering. pages 470–479. ACM, 2007.