contributed articles

Strong ties connect individuals in the same community; weak ties connect individuals in different communities.

BY PASQUALE DE MEO, EMILIO FERRARA, GIACOMO FIUMARA, AND ALESSANDRO PROVETTI

On Facebook, Most Ties Are Weak

PERVASIVE SOCIO-TECHNICAL NETWORKS bring new conceptual and technological challenges to developers and users alike. A central research theme is evaluation of the intensity of relations linking users and how they facilitate communication and the spread of information. These aspects of human relationships have been studied extensively in the social sciences under the framework of the "strength of weak ties" theory proposed by Mark Granovetter.[13] Some research has considered whether that theory can be extended to online social networks like Facebook, suggesting interaction data can be used to predict the strength of ties. The approaches being used require handling user-generated data that is often not publicly available due to privacy concerns.

Here, we propose an alternative definition of weak and strong ties that requires knowledge of only the topology of the social network (such as who is a friend of whom on Facebook), relying on the fact that online

social networks, or OSNs, tend to fragment into communities. We thus suggest classifying as weak ties those edges linking individuals belonging to different communities and strong ties as those connecting users in the same community. We tested this definition on a large network representing part of the Facebook social graph and studied how weak and strong ties affect the information-diffusion process. Our findings suggest individuals in OSNs self-organize to create well-connected communities, while weak ties yield cohesion and optimize the coverage of information spread.

Analysis and understanding of OSNs like Facebook takes a theoretical foundation in social network analysis from Kleinberg.[15] However, studying a real OSN poses computer science challenges, given the size, distribution, and organization (such as privacy and visibility rules) of data available to regular OSN subscribers.[1]

In this context, analysis of large subsets of OSNs should generate a series of statistically robust measurements that are the basis of understanding OSN structure and evolution. Indeed, aggregate measures are very valuable in data management, privacy management, and online marketing. A challenging problem is how to evaluate the intensity of relations that bind users and determine how they facilitate the spread of information. These aspects of user connectedness and social communication have been studied in social sciences before, notably through Granovetter's theory.[13]

» key insights

- **Online social networks can facilitate the growth of a friendship network consisting of distant communities; severing weak ties has a strong effect on the connectivity of the overall network.**

- **We redefine the strong-versus-weak ties relationship to reflect the structure and evolution of massive online social networks, including Facebook.**

- **Interesting properties emerge when users self-organize into communities within online social networks.**

Weak ties are connections between individuals belonging to distant areas of the social graph, or the ones that happen to have most of their relationships in different national, linguistic, age, or common-experience groups. Weak ties are a powerful way to transfer information across large social distances and to wide segments of the population. Strong ties are contacts between trusted/known persons (such as family ties and close friendships).

Are Granovetter's weak ties also found in OSNs like Facebook in the form he envisioned, or connections between individuals belonging to different areas of the social graph? Answering is difficult for at least two reasons: Facebook is organized mainly around the recording of just one type of relationship—friendship—implying Facebook friendship captures (and compresses) several degrees and nuances of human relationships dif-ficult to separate and characterize through analysis of online data. And second, as Facebook grows in size and complexity, its friendship network grows denser, not sparser.[1]

As OSNs become increasingly inter-connected, testing Granovetter's theory poses scalability challenges. Early research, as in Gilbert and Karahalios,[9] took a supervised approach, where a panel of Facebook users was asked to assess the strength of their own

friendship ties. Large-scale studies of Granovetter's theory, as in Gilbert and Karahalios,[9] would arguably be difficult to conduct, given the size of today's OSNs. Other approaches, notably Bakshy et al.,[2] access Facebook data on user activities and compute tie strength as a function of type and frequency of user interactions. However, a cutoff threshold is required to distinguish strong ties from weak ties, and the tuning of that threshold has a crucial effect on identification of weak ties.

Here, we propose a new definition of weak ties rooted in analysis of large OSNs and in light of related computational challenges. We first identify communities within the network, then classify as weak ties those edges that connect users located in different communities; strong ties are those edges between users in the same community. Identifying weak ties through our definition is quick due to the efficiency of recent algorithms for finding communities in large networks[8] and robust because no threshold has to be defined. We performed extensive experimental analysis on a public dataset on Facebook friendship collected and released by Gjoka et al.[10] and a null-model comparison against randomly generated graphs. We deployed two well-known community-detection algorithms: the Louvain Method (LM)[3] and Infomap.[20]

Our analysis of the experimental results produced the following insights:

Independent of algorithm. The weak ties discovered through LM tend to coincide with those found through Infomap; our definition of weak ties is thus largely independent of the choice of community-detection algorithm;

Weak outnumber strong. Our community-defined weak ties outnumber strong ties;

More frequent. Weak ties occur more frequently in communities of small size; and

Spread of information. Weak ties identified through our approach play a crucial role in spreading information over a network, and their removal reduces the portion of the network that can be reached through information diffusion.

Weak and Strong Ties

Mark Granovetter introduced his now-classic definition:[13] "The strength of a

We suggest classifying as weak ties those edges linking individuals belonging to different communities and strong ties as those connecting users in the same community.

tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services that characterize the tie." It introduces important features of social ties like intensity of connection, or frequency of contacts, and reciprocity. Granovetter considered identification of strong and weak ties through topological information related to the social network structure. He thus introduced the concept of "bridge": "A bridge is a line in a network which provides the only path between two points. Since, in general, each person has a great many contacts, a bridge between A and B provides the only route along which information or influence can flow from any contact of A to any contact of B." Note, all bridges are weak ties. Unfortunately, with OSNs, the definition of bridge is restrictive and unsuitable. The well-known "small world effect," or presence of short paths connecting a pair of vertices, and the "scale-free degree distribution," or presence of "hubs" maintaining the whole network, make it unlikely that an edge could be found whose deletion would completely disconnect its terminal vertices.

To adapt Granovetter's definition to OSNs, we could redefine a "shortcut bridge" as a link connecting a pair of vertices whose deletion would cause an increase in the distance (defined as the length of the shortest path linking two vertices) between them. However, this otherwise sensible definition is more controversial than it might seem. First, it depends on the notion of shortest path, but, unfortunately, identification of all-pairs shortest paths is computationally unfeasible, even on networks of modest size. Second, even the alternative definition of distance, or "weighted path" (computed as the sum of the weights assigned to each connection being used), would remain computationally prohibitive for OSNs. Our goal here is to explore a novel and computationally feasible definition of weak ties that suits analysis of large-scale OSNs. Instead of a relational definition based on the intensity of the interactions between two users, we propose a community-based definition. We define weak ties

as those edges that, after dividing up the network into communities (thus obtaining so-called "community structure"), connect vertices belonging to different communities. Finally, we classify intra-community edges as strong ties.

One of the most important features of our weak ties is that those that are bridges (in the Granovetter sense) create more, and shorter, paths; their deletion would indeed be more disruptive than the removal of a strong tie, from a community-structure perspective. This may even be the reason weak ties (albeit defined in a slightly different fashion) have proved effective in the diffusion of information.[5,21]

Benefits of our definition. Our definition of weak tie has four appealing features:

Weaker than Granovetter. The fact that vertices linked by a weak tie belong to different communities does not imply the edge between them is a bridge; its deletion may not disconnect its vertices, and, in practice, almost never does;

Based on topological information. It enables weak/strong classification on the basis of only topological information. Zhao et al.[21] used topological information but only locally, on the neighbors of the two terminal vertices, whereas our definition handles global information, as it relies on the partitioning of the whole network;

Binary. It labels each edge in the network as either weak or strong; as a consequence, there is no need to need to "tune" any threshold, below which edges are classified as weak, the upshot being we cannot compare two edges on the basis of their strength; and

Accuracy of community discovery. Our experiments show our definition of weak tie is robust with respect to the choice of a particular community-detection algorithm.[8]

Related Approaches
Several research projects have examined the strength of weak ties in terms of nontopological information. A nontopological approach to determining degrees of strength or weakness requires researchers to choose some measurable variables by which the strength of the relation binding two users can be deduced. In such a scenario, weak ties are intended in a relational sense; that is, they connect acquaintances who do not interact frequently and therefore do not influence each another strongly. For "offline" social networks, Marsden and Campbell[17] identified some variables (such as communication reciprocity and the presence of at least one mutual friend) as indicators of a weak tie.

In the context of the social Web, Gilbert and Karahalios[9] and Petroczi et al.[19] studied small OSNs—35 participants and 56 participants, respectively—assigning weights to the relationships on the basis of several measures of strength (such as intimacy/closeness, reciprocity, and sociability/conviviality) assessed through direct questions to participants. Gilbert and Karahalios[9] extended Marsden-Campbell's method to Facebook by identifying as many as 74 variables as potential predictors of strength. They then modeled strength as a linear combination of the variables, computing the weights by means of a variant of ordinary least-square regression. To validate their model, Gilbert and Karahalios[9] recruited 35 users and asked them to rate their Facebook friends. They achieved accuracy of approximately 85%, but their performance is difficult to replicate on large OSN fragments. Their method requires collecting a large amount of information on user behavior; due to privacy concerns and the limitation on use of proprietary data, most such required data is unlikely to be available for academic studies.

Zhao et al.[21] formalized the strength of ties with a weight w_{ij} assigned to the edge connecting vertices i and j as follows

$$w_{ij} = \frac{c_{ij}}{k_i + k_j - 2 - c_{ij}}$$

where k_i and k_j are the degrees of i and j and c_{ij} are the numbers of their mutual acquaintances. Numerical simulations by Zhao et al.[21] showed that by gradually deleting ties with lower w values, information coverage drops sharply.

Zhao et al.'s results corroborate the hypothesis that weak ties are in fact the key to information diffusion. A 2012 paper on weak ties in Facebook due to Bakshey et al.[2] considered four parameters to measure the strength of a connection: frequency of private messages; frequency of public comments left on each other's posts; number of times they jointly appear in a photo; and number of times they jointly commented on a third-party post.

Unlike us, Bakshy et al.[2] did not consider, at least not explicitly, the topological aspects of Facebook social connections; moreover, their analysis requires access to proprietary data and records of user activity, so the methodologies are not easily compared. Our work is more like Blondel et al.[2] and Zhao[21] in that weak ties are understood to be useful connectors favoring the spread of information and was (as in Blondel et al.[2]) tested on Facebook.

However, one difference emerges: Blondel et al. and Zhao et al. both assigned scores to ties and classified them according to a threshold value. In contrast, we classify ties as weak or strong, depending on whether or not they connect vertices located in different communities; our classification scheme is binary and does not use scores and threshold, which may be difficult to set up and tune properly. Another approach is due to Grabowicz et al.,[12] who, like us, used information about network topology to identify weak ties. However, they focused on Twitter, which can be modeled as a directed network where user relationships are mostly asymmetric. The Twitter "retweet" and "mention" functions have a major effect on identification of weak ties; the studies are thus not easily compared.

Results
The results of the tests we carried out highlight the pros and cons of our definition of a weak tie. For the preliminary community-detection phase we deployed two popular algorithms: LM[3] and Infomap.[20] We considered two testbeds, comparing them with a fragment of the Facebook network collected by Gjoka et al.[10] with 957,000 users and 58.4 million friendship connections[a] and "null-model" networks consisting of Erdös-Rényi random graphs with up to 2,048 vertices; we varied the

---

a   For a complete description of the dataset, see supplementary material at http://informatica.unime.it/weak-ties/.

Figure 1. The complementary cumulative distribution function associated with (a) weak ties and (b) strong ties in our Facebook dataset.
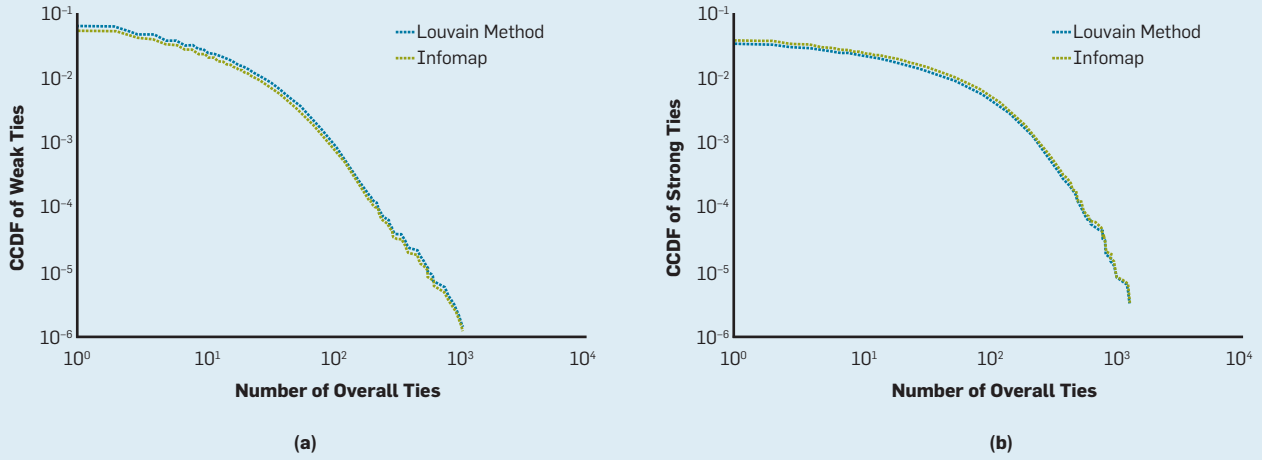

(a)

(b)

Figure 2. The R_{avg} ratio for Erdös-Rényi random graphs clustered with (a) LM and (b) Infomap; instances with $|V| = 128, 512, 1,024,$ and $4,096$.
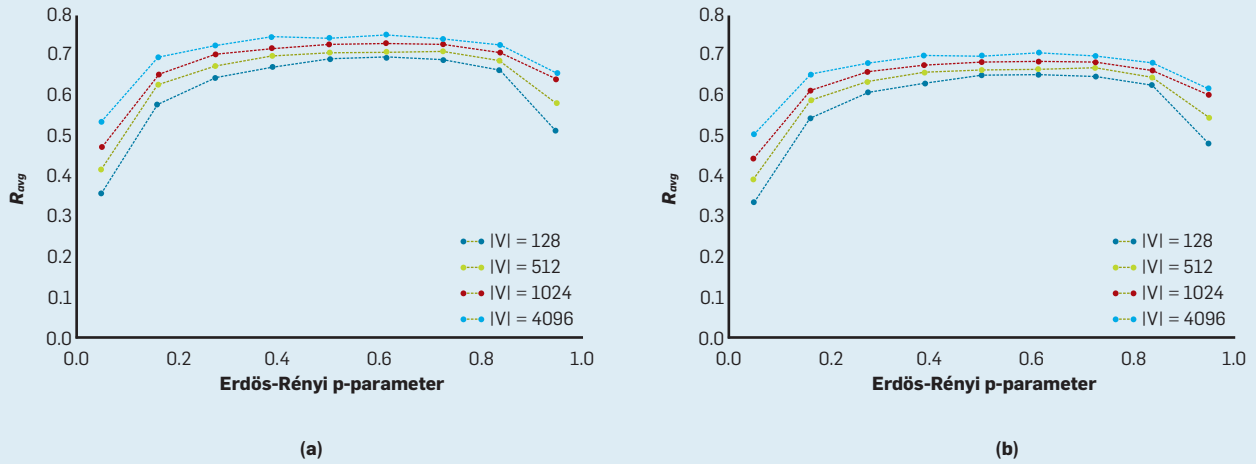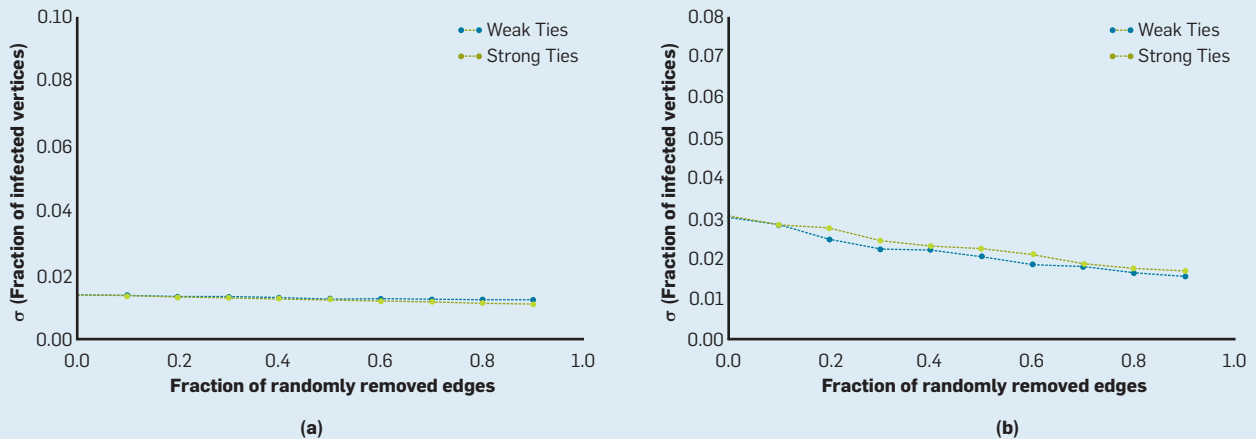

(a)

(b)

Figure 3. Coverage in graphs with different values of τ: (a) with Erdös-Rényi's random graphs ($|V| = 512$, plink = 0.05, pinf = 0.01) when τ ranges from 0 to 0.9; (b) with the Facebook dataset generated with τ ranging from 0 to 0.9. Charts (a) and (b) report results for the LM.

(a)

(b)

probability of having an edge between an arbitrary pair of vertices uniformly from 0.05 to 0.95, with a 0.05 step.

Robustness of definition. As a preliminary experiment, we studied the robustness of our definition of weak ties with respect to the method one might adopt for finding communities. Different community-detection methods might produce different results, albeit slightly, hence our weak/strong ties classification could become dependent on the community-detection method. We ran both LM and Infomap on our Facebook sample, comparing the community structures found by the two algorithms by applying the normalized mutual information (NMI). We found an NMI of approximately 0.9, informing us that the level of disagreement between the two algorithms is quite low.[b]

We conclude that on Facebook our definition of weak ties is robust because it does not depend on a specific community-detection algorithm. However, note that the communities discovered by LM and Infomap might differ in subtle ways yet still be characterized by a high NMI. We next offer a detailed, graphical comparison of the results we achieved by applying LM and Infomap to better understand the practical consequences of adopting a particular community-detection algorithm.

Strong and weak ties on Facebook. Here, we consider the distribution of weak and strong ties in Facebook. We computed the "complementary cumulative distribution function," or CCDF, of weak (as opposed to strong) ties, or the probability of finding a vertex with more than k weak (as opposed to strong) ties (see Figure 1).

The CCDF associated with strong ties decreases quicker than that for weak ones. At $k = 4$, we see a tipping point after which weak ties quickly outnumber strong ties; that is, the latter are much less frequent than the former in higher-degree vertices.

The results of our experiments agree fairly well with Granovetter's "vi-

---

b   The definition of NMI is included in the supplementary material (see footnote a); if NMI, which ranges between 0 and 1, approaches 1, then the communities found by the two algorithms can be viewed as coinciding.

Our findings suggest individuals in OSNs self-organize to create well-connected communities, while weak ties yield cohesion and optimize the coverage of information spread.

sion" of weak ties. Sociological theories (such as triadic closure,[13] cognitive balance,[14] and homophily[18]) suggest individuals tend to aggregate in small communities.

We thus confirm that the intensity of human relations is very tight within small groups but decreases for individuals belonging to distant communities. Most connections are weak in the sense of Granovetter, with few contacts and infrequent interactions.

Weak and strong ties in random graphs. Here, we offer a comparative analysis of weak/strong ties distribution in Erdös-Rényi random graphs; that is, we check whether the results we obtained on Facebook data still hold on graphs where the structure is known in advance and lacks, by construction, a clear community structure. We computed the ratio R_{avg} of the number of weak ties to the total number of ties; Figure 2 outlines R_{avg} for different values of $|V|$ and when the probability p_{link} of having an edge between an arbitrary pair of vertices varies uniformly from 0.05 to 0.95.

We thus derive several insights:

Weak outnumber strong. R_{avg} is always greater than 0.6; that is, weak ties still outnumber strong ties, even in randomly generated graphs;

Stable and independent. R_{avg} is relatively stable and independent of p_{link}; that is, the "sparseness" of G has a limited effect on the number of weak ties; and

Limited effect. $|V|$ has a limited effect on R_{avg}; when $|V|$ goes from 128 to 4,096, or increases by a factor of 32, R_{avg} increases by only 17.14%.

Weak ties in information diffusion. We also studied how weak ties influence the information-diffusion process, clarifying the connection between our definition and Granovetter's, since we view weak ties as providing specific links between individuals who would otherwise remain disconnected, as they belong to distant areas of the social graph.

Granovetter's weak ties should play a central role in the spread of information. But how do our weak ties convey information?

We applied the Independent Cascade Model (ICM)[11] to simulate information propagation over a network, again taking the Facebook dataset

and Erdös-Rényi's random graphs for comparison. In ICM, a vertex v_0 is selected uniformly at random to forward a message to its neighbors, with probability equal to p_{inf} (infection probability).

Each "infected" vertex can, in turn, recursively propagate the message to its neighbors. Leskovec et al.[16] reported reasonable values of p_{inf} are 0.01, 0.02, and 0.03.

To generate statistically significant results, we selected the vertex v_0 to start from 250 times; at each selection of v_0, we simulated the propagation of a message, measuring the coverage σ defined as the ratio of the number of vertices receiving a message (infected vertices) to the total number of vertices. We repeated the experiment by progressively (and randomly) deleting a fraction τ of weak ties.

In our simulation, τ ranged from 0.1 to 0.9, and, for each value, we computed the corresponding coverage. The whole procedure was repeated separately for weak and strong ties. Figure 3 and in the supplementary material (http://informatica.unime. it/weak-ties/) show the values of σ obtained by applying the LM and Infomap on Erdös-Rényi's random graphs, respectively. Here, we consider a graph with $|V|$ = 512 vertices, p_{link} = 0.05 and p_{linf} = 0.03.

Note the gradual deletion of weak ties yields a decrease of σ; on average, the greatest decrease of σ is 11.98%, with an average decrease equal to 5.71%, with standard deviation of 3.4%. We found the same behavior when removing strong ties, but the decrease was less marked. We observed that if $|V|$ increases, the coverage σ also increases. The values of σ associated with our Facebook dataset are reported in Figure 3b and in the supplementary material.

Again, for a fixed value of τ, removing weak ties yields a decrease of σ more marked than the removal of strong ties. The greatest decrease of σ we observed was approximately 14.31%, with an average decrease of 6.26%, with standard deviation of 3.1%. These values are always greater than those for Erdös-Rényi's random graphs, proving the removal of weak ties is more significant in real OSNs than in networks not equipped with a meaningful community structure. We thus conclude that our definition of weak ties captures Granovetter's idea, in that deleting weak ties decreases/obstructs the flow of information much more effectively than removing strong ties.

Conclusion

We have presented a novel definition of weak ties designed for OSNs like Facebook based on the community structure of the network itself. The experiments we conducted on a Facebook sample of 957,000 users and randomly generated graphs highlight the role and importance of weak ties. We characterized the overall statistical distribution of weak ties as a function of the size of a community and its density. We studied their role in information-diffusion processes, with results suggesting a connection between our definition of weak ties for OSNs and Mark Granovetter's original definition.

Even though several recent research projects have focused on Facebook's social graph,[1,4] community structure,[7] and weak ties per se,[2] our community-based definition of weak ties better fits Facebook and similarly large (and dense) OSNs.

As for future work, we hope to follow up with two more projects: The first concerns investigating the applicability of network-weighting strategies, so the strength of ties can be computed according to a given rationale (such as the ability of each link to spread information); we intend to adopt a novel method of weighting edges suited for OSNs we devised in De Meo et al.[6] The second concerns analysis of geographical data related to Facebook users; due to the merging of different graphs (such as social and geographical), we expect to gain additional insight into the role of physical vs. virtual distances.

Acknowledgments

References
1. Backstrom, L., Boldi, B., Rosa, M., Ugander, J., and Vigna, S. Four degrees of separation. In *Proceedings of the ACM Web Science Conference* (Evanston, IL, June). ACM Press, New York, 2012, 33–42.
2. Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. The role of social networks in information diffusion. In *Proceedings of the World Wide Web Conference* (Lyon, France). ACM Press, New York, 2012, 519–528.
3. Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* (Oct. 2008).
4. Catanese, S., De Meo, P., Ferrara, E., Fiumara, G., and Provetti, A. Crawling Facebook for social network analysis purposes. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics* (Sogndal, Norway, May 25–27). ACM Press, New York, 52:1–52:8.
5. Centola, D. The spread of behavior in an online social network experiment. *Science 329* (Sept. 2010), 1,194–1,197.
6. De Meo, P., Ferrara, E., Fiumara, G., and Ricciardello, A. A novel measure of edge centrality in social networks. *Knowledge-Based Systems 30* (June 2012), 136–150.
7. Ferrara, E. A Large-scale community structure analysis in Facebook. *EPJ Data Science 1*, 1 (Nov. 2012), 1–30.
8. Fortunato, S. Community detection in graphs. *Physics Reports 486* (Feb. 2010), 75–174.
9. Gilbert, E. and Karahalios, K. Predicting tie strength with social media. In *Proceedings of the 27th International Conference on Human factors in Computing Systems* (Boston, Apr.). ACM Press, New York, 2009, 211–220.
10. Gjoka, M., Kurant, M., Butts, C.T., and Markopoulou, A. Practical recommendations on crawling online social networks. *IEEE Journal on Selected Areas in Communications 29*, 9 (Oct. 2011), 1,872–1,892.
11. Goldenberg, J., Libai, B., and Muller, E. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters 12*, 3 (2001), 211–223.
12. Grabowicz, P.A., Ramasco, J.J., Moro, E., Pujol, J.M., and Eguiluz, V.M. Social features of online networks: The strength of intermediary ties in online social media. *PLOS ONE 7*, 1 (Jan. 2012).
13. Granovetter, M.S. The strength of weak ties. *American Journal of Sociology 78*, 6 (May 1973), 1360–1380.
14. Heider, F. *The Psychology of Interpersonal Relations*. Lawrence Erlbaum, New York, 1982.
15. Kleinberg, J. The convergence of social and technological networks. *Commun. ACM 51*, 11 (Nov. 2008), 66–72.
16. Leskovec, J., Adamic, L., and Huberman, B. The dynamics of viral marketing. *ACM Transactions on the Web 1*, 1 (May 2007).
17. Marsden, P.V. and Campbell, K.E. Measuring tie strength. *Social Forces 63*, 2 (Dec. 1984), 482–501.
18. McPherson, M., Smith-Lovin, L., and Cook, J.M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*, 1 (Aug. 2001), 415–444.
19. Petroczi, A., Nepusz, T., and Bazs, F. Measuring tie-strength in virtual social networks. *Connections 27*, 2 (2006), 39–52.
20. Rosvall, M. and Bergstrom, C.T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences 105*, 4 (Jan. 2008), 1,118–1,123.
21. Zhao, J., Wu, J., and Xu, K. Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E 82*, 1 (July 2010).

Pasquale De Meo (pdemeo@unime.it) is an assistant professor in the Department of Ancient and Modern Civilizations at the University of Messina, Italy.

Emilio Ferrara (ferrarae@indiana.edu) is a postdoctoral researcher in the School of Informatics and Computing at Indiana University, Bloomington, IN.

Giacomo Fiumara (gfiumara@unime.it) is an assistant professor in the Department of Mathematics and Informatics at the University of Messina, Italy.

Alessandro Provetti (ale@unime.it) is an associate professor in the Department of Mathematics and Informatics at the University of Messina, Italy.