Disengagement Social Contagion: Assessing Network Vulnerability under Node Departures

Fragkiskos D. Malliaros[1], Michalis Vazirgiannis[1,2]
[1]École Polytechnique, France
[2]Athens University of Economics and Business, Greece
{fmalliaros, mvazirg}@lix.polytechnique.fr

Keywords

Social Engagement; Network Vulnerability; Cascading Behavior

1. INTRODUCTION AND PRELIMINARIES

Understanding the properties and dynamics of social networks, arising from online social networking and social media platforms, is an interesting task with plenty of applications in both the Web and social sciences. Typically, the structure of social networks is not static but is governed by an increased level of evolution. Users decide to join in online communities for various reasons (e.g., create new friendship relationships), that mostly express means of interaction among individuals in the social web.

Furthermore, it is expected that some users may decide to leave the network, or in general to stop being active in the activities of their community. This phenomenon, also known as churn or attrition, has been an important topic in the business domain. The key point here is that this decision can affect the decisions of their neighbors in the social graph that, in their turn, may decide to depart as well. That way, the departure of a single user (node) can become a *disengagement epidemic*, forming a *cascade* of potential individual departures that may have consequences to the overall structure of the network. Being able to model and analyze such phenomena in real social networks is an important task, since they are related to the vulnerability of these social interaction systems under node departures.

The problem of robustness (or vulnerability) assessment in real networks has been extensively studied by different research communities, including sociology, statistical physics and computer science. The observation of the power-law degree distribution in real networks [1] was the basis for several works (e.g., [2, 6]), which shown that real networks are robust against random failures but vulnerable under attacks to high degree nodes.

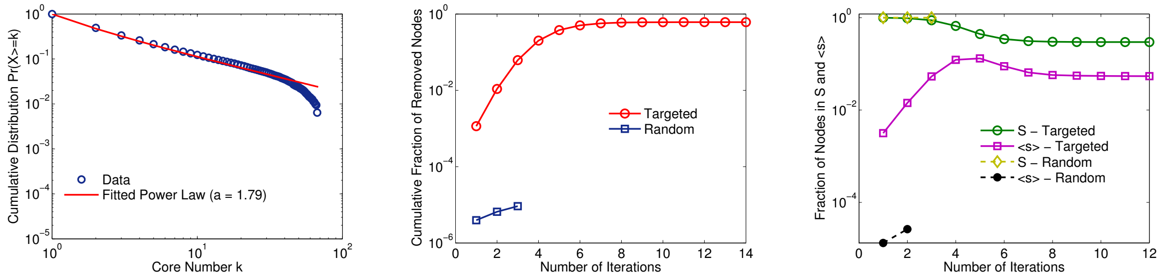
However, in the case of social networks, instead of degree-based failures and attacks, users decide to depart from or stay in the network based on their own *engagement level*. Recent studies about the departure dynamics in social networks suggest that the engagement level of nodes, using topological features of the network, are not accurately described by the node degree [3, 5, 4] and therefore, well-known degree-based types of robustness assessment may not accurately capture this feature of social networks. Under these settings, it has been proposed (based on theoretical and experimental results) that the engagement level of a node $i \in V$ can be captured by the *core number* c_i, as produced by the k-core decomposition [3, 5, 4]. A subgraph H of G is defined to be a k-core of G, if it is a maximal connected subgraph of G, in which all nodes have degree at least k. Then, node i has core number $c_i = k$, if it belongs to a k-core but not to any $(k+1)$-core. That way, nodes with high core number c_i show better engagement properties and therefore, it is less probable to depart from the graph. The idea behind this modeling approach is based on the following game theoretic concept: nodes that want to gain an explicit benefit by remaining engaged in the network, they should align their decision (leave or stay) with the one of their neighbors. Furthermore, the core number c_i combines in a meaningful way two intuitive requirements for high engagement namely, node degree d_i and the level of interaction among neighborhood nodes.

The goal of this paper is to introduce and study a novel problem of vulnerability assessment in social networks, under cascades caused by node departures based on their engagement level. Initially, we propose the *Cascading Departure (CasD) model*, a k-core decomposition based model to capture the cascading (epidemic) disengagement effect due to the departure of a node. Then, combining the property of power-law core number distribution observed in social networks (see Fig. 1 (a)) with the proposed Cascading Departure model, we introduce a new problem of vulnerability assessment in social graphs based on cascades triggered by *random* and *targeted* node departures based on the core-number. We have performed and present preliminary experiments on real graphs, and our key observation is that online social networks are extremely robust under cascades started by random departures of nodes; however, they are highly vulnerable under cascades caused by targeted departures of nodes with high engagement level.

2. GRAPH VULNERABILITY UNDER NODE DEPARTURES

Initially, we introduce the proposed Cascading Departure (CasD) model. As we have already discussed, the departure of a node can cause direct effects in its neighborhood, in the sense that some of the friends in the social graph may also decide to depart – leading to an epidemic of disengagement. Let $G = (V, E)$ be the undirected graph that models a social network. Suppose that a node $v \in V$ decides to depart (next we will describe how this node can be selected, namely randomly or targeted) and let $\tilde{V} = V \setminus \{v\}$ be the remaining node set. At each time step of the model, two points need to be specified: (i) how to determine if a departure affects a neighborhood node, and (ii) how the affected by the cascade nodes decide to depart.

To address these points, we capitalize on the relationship between the engagement property and the core number, as described earlier [3, 5, 4]. Let c_i and $\tilde{c}_i$ be the core numbers of a node i before and after the departure respectively. Each node $i \in V$ has an engagement level – that can be captured by the core number c_i – which expresses the incentive of the node to remain in the graph (or inversely, to depart). Thus, we consider that the nodes that are affected by a departure are those which their core number c_i has changed after the departure of node v. We know that after the dele-



(a) Core number distribution (b) Fraction of removed nodes (c) Network fragmentation

Figure 1: Results on the EPINIONS social network. (a) Cumulative core number distribution. Observe that the distribution is heavy-tailed. (b) Cumulative fraction of removed (affected) nodes per iteration of the CasD model (ten runs), under random and targeted node departures based on their engagement level. (c) Fraction of nodes (logarithmic scale) in the largest connected component S and in the rest isolated connected components $\langle s \rangle$ per iteration of the CasD model (ten runs), under random and targeted departures.

tion of a node, the core number of each node $\tilde{c}_i$ in the graph, can either be reduced by one or remain the same. Thus, if $\tilde{c}_i < c_i$, node i is characterized as affected by the cascade. Furthermore, in order to specify if an affected node $i \in \tilde{V}$ will finally depart, we consider that the probability of departure should be inversely related to the core number $\tilde{c}_i$ (nodes with lower core number are more probable to leave). Let $\tilde{c}_i^{\text{norm}} = \frac{\tilde{c}_i - \min(\tilde{\mathbf{c}})}{\max(\tilde{\mathbf{c}}) - \min(\tilde{\mathbf{c}})}, \forall i \in \tilde{V}$ be the normalized core number of node i in the range of $[0, 1]$, where $\tilde{\mathbf{c}} = [\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_{|\tilde{V}|}]$ is a vector that contains the core numbers of each node, and $\min(\cdot), \max(\cdot)$ are functions that return the minimum and maximum element of a vector. Then, node i is removed from the graph with probability $\Pr\left(\tilde{V} = \tilde{V} \setminus \{i\}\right) = 1 - \tilde{c}_i^{\text{norm}}$. This process is repeated as long as nodes continue to be affected by departures during the previous time step. In order to quantify the disengagement effect of a departure, we keep track of the set of removed nodes R. As we will present shortly, the size of this set depends heavily on how the initial node v of the cascade is selected.

Note that, in contrast to other degree-based models which mostly consider features of technological networks (like the Internet) that are responsible for functional *errors* (e.g., [6, 1]), the proposed model naturally captures the social component of an epidemic process in social networks, in the sense that the decision of individuals can potentially be affected by the decisions of other individuals within their social environment.

The CasD model is heavily based on the engagement level of each node, as captured by the core number. We have examined the core number distribution of several social networks, and in Fig. 1 (a) we depict the results for the EPINIONS social network[1]. As we can observe, core numbers follow a heavy-tailed distribution, where most of the nodes demonstrate small core number and therefore their engagement level tends to be low. On the other hand, only a few nodes have high core number and these nodes can be considered as the most engaged ones. That way, if we randomly select a node, this node is more probable to have low core number due to the skewness of the distribution.

Based on these points, we define two different strategies of node departures, i.e., how to select a single node $v \in V$ that will depart first and trigger a cascade: (i) *random departure*: a randomly selected node leaves the graph; (ii) *targeted departure*: a node selected among the ones with the highest core number decides to depart. The first strategy simulates what is more probable to occur in practice (nodes with lower engagement are more probable to depart). The strategy of targeted departures captures the case in which a node, although it does not have incentive to leave (as expressed by a high core number), it finally departs due to external factors (such as an adversary that motivates a user to disengage from the activities of the network). To study the dynamics of these strategies, we apply the CasD model selecting accordingly the initial node v. To assess the vulnerability of social networks, we examine the total fraction of removed nodes during the execution of the model (i.e., for the time steps that the epidemic is spreading), and the fragmentation of the graph as captured by the sizes of the largest connected component S and the rest isolated components $\langle s \rangle$. As we can observe from Fig. 1 (b), in the case where the initial node is selected randomly, the fraction of affected nodes is extremely small and the epidemic typically dies out early. On the other hand, targeted departures have the potential to affect a large portion of the graph, typically more than 50% of the nodes, and this behavior is persistent for all the datasets that we have examined. Additionally, as Fig. 1 (c) depicts, the fragmentation of graph is much more intense in the case of a cascade triggered by a targeted departure.

These results indicate that social networks are extremely robust under cascades triggered by the departure of randomly selected nodes, but they tend to be highly vulnerable in cascades caused by targeted departures of nodes with high engagement level. This suggests an additional *robust-yet-fragile* property of networks with heterogeneous structural characteristics. We consider that the proposed problem of vulnerability assessment is more close to what really occurs in social networks, and suggests several directions for future work. One possible direction could be to further validate the predictive cascade capabilities of the model by examining departure (or inactivity) traces of real networks. The CasD model can be thought of as an epidemic process and therefore a more thorough theoretical analysis of its properties is also an interesting direction.

3. REFERENCES

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.

[2] R. Albert, H. Jeong, and A. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794), 2000.

[3] K. Bhawalkar, J. Kleinberg, K. Lewi, T. Roughgarden, and A. Sharma. Preventing unraveling in social networks: the anchored k-core problem. In *ICALP*, pages 440–451. 2011.

[4] D. Garcia, P. Mavrodiev, and F. Schweitzer. Social resilience in online communities: The autopsy of friendster. In *COSN*, pages 39–50, 2013.

[5] F. D. Malliaros and M. Vazirgiannis. To stay or not to stay: Modeling engagement dynamics in social graphs. In *CIKM*, pages 469–478, 2013.

[6] A. E. Motter and Y.-C. Lai. Cascade-based attacks on complex networks. *Phys Rev E*, 66:065102, 2002.

[1] Who-trusts-whom social network, extracted from the product review website `www.epinions.com`. The graph has $75,877$ nodes and $405,739$ edges. Data from: `snap.stanford.edu`.