Social Leaders in Graphs

V. Blondel[1], C. de Kerchove[2], E. Huens[3] and P. Van Dooren[4]

[1] Université catholique de Louvain (UCL), Department of Applied Mathematics, Avenue Georges Lemaître, 4 B-1348 Louvain-la-Neuve Belgium
`blondel@inma.ucl.ac.be`
[2] UCL `dekerchove@inma.ucl.ac.be`
[3] UCL `huens@inma.ucl.ac.be`
[4] UCL `vandooren@inma.ucl.ac.be`

We introduce the definition of Social Leader that gives a local centrality measure for each node in a graph. A node u is a Social Leader if the number of cycles of length 3 passing through u is greater than the corresponding number for its neighbors. This concept is used to visualize large graphs, identify influent agents in social networks (word of mouth effect) and find communities.

1 Introduction

The principle of social leader appears in the context of social networks [1]. Emails, telephone calls, friendship relations, coauthors, ... are examples of social networks that are widely explored in the literature. Nonnegative matrices techniques allow us to represent them and to study dynamical systems evolving on a social network. This is essential for viral marketing where we need to identify agents that are influent in the network in order to speed up the word of mouth effect. Influent persons are not only well connected with many contacts but their contacts also have a large probability to know each other. This leads to the concept of a social leader.

Formally, let $G(N, E)$ be any undirected graph where N and E are respectively its set of nodes and edges. We define first the social degree of a node u that is used to determine the social leaders.

Definition 1. *The social degree of node u in G, $SD(u)$, is the number of cycles of length 3 passing through u.*

We point out that the sum of all SDs is equal to 3 times the number of triangles in the graph, i.e. the number of cycles of length 3. This number is high for social networks since we have transitivity of friendship. A first condition to be a social leader is to have a SD different from 0, in other words to belong at least to a trio of friends. The second condition is based on a local maximum: a social leader has a SD greater or equal to the SDs of his contacts, see Figure 1.

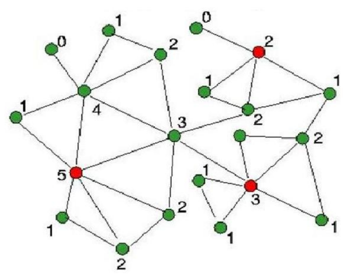

Fig. 1. A social degrees distribution with three social leaders.

We point out that two social leaders cannot be neighbors except when they have the same SD. A typical example is a clique: every node is connected to every node and then all nodes have the same SD. However the adjacency of connected social leaders, as in cliques, allows us to have a separation between groups of social leaders and isolated social leaders. The aggregation of a graph around these two types of nodes is discussed in the section of visualization of large graphs. Social leaders can be viewed as an optimal meeting point for their neighborhood and have interesting applications as we will see in the next section. Moreover the calculation of SDs and social leaders in a graph G has a linear complexity $O(m \cdot d_{max})$, where m is the number of edges in G and d_{max} the largest number of contacts in G. On the other hand the SD of a social leader could be only slightly greater than the SDs of its neighborhood. Such social leaders can be dethroned easily by small perturbations on edges or nodes. That leads us to consider the sensitivity of social leaders and also to extend previous definitions for other types of graphs like evolving networks (last section).

2 Applications

2.1 Visualization of large graphs

Several algorithms for visualizing large graphs consider a repulsive force between every pair of nodes and an attractive force between every pair of connected nodes. Then they calculate the equilibrium of these forces and project this in the plane [2, 3]. These methods provide good results, but are expensive for graphs with millions of nodes. We studied a graph of telephone calls where nodes are the customers and the edges are calls between them. That graph had 2 millions nodes and 10 millions edges, and in order to visualize it, we aggregated the graph around its social leaders for several recursive levels until reaching a graph of about 60.000 nodes. The aggregation process of $G(N, E)$ considers only the social leaders and adds an edge between two social leaders if there exists a path of length at most 3 in the graph G, see Figure 2. More details will be given in the full version.

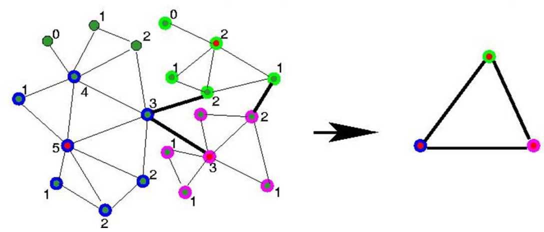

Fig. 2. Aggregation from left to right.

From the reduced graph, we use the algorithm in [3] that calculates the equilibrium of the forces in the graph. Figure 3 shows that visualization with flemish, inhabitant of Brussels and walloon users.

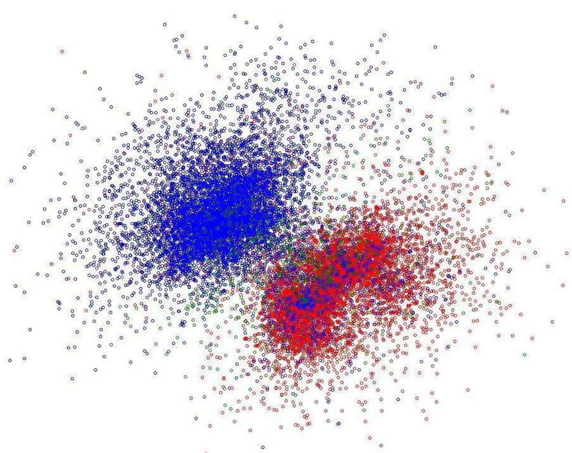

Fig. 3. Blue: flemish users; Green: inhabitant of Brussels; Red: walloon users.

2.2 Identification of leaders in social networks

The telephone calls graph is a typical example of social network. In viral marketing, we identify users (i.e. nodes) that play an essential role in the word of mouth effect. Our results show that social leaders that have the MMS technology (i.e. the possibility to send image files) are two times more successful in infecting their community with that technology than the customers that have a similar profile but are not social leaders. The graph of telephone calls evolves over time and exposes a complex dynamical system with phenomena of pressure and influence between connected nodes. Data from mobile phone companies gives the opportunity to have relevant benchmarks for such networks.

2.3 Finding communities

Several greedy algorithms for clustering start with a set of nodes called the seed. Then they add edges and nodes from that seed according to different rules, see [4]. Since social leaders are locally well-connected, they should be good candidates to initialize such greedy algorithms. Other applications can be found in multi-level problems where the coarse grid is represented by social leaders. More details will be given in the full version.

3 Extensions

We introduced the concept of social leaders for static and undirected graphs. But, as mentioned above, the graph of telephone calls is dynamic, meaning that nodes and edges are changing over time. Moreover that graph is directed and weighted since each call has a direction and a duration. The determination of social leaders for weighted digraphs is straightforward if we redefine the SDs for such graphs. Several heuristics can be applied. For example, given a digraph, there are four categories of cycles of length 3 represented in Figure 4. So we can assign a different score to each cycle. The SD of a node u is then the sum of scores of the cycles passing through u. Similar ideas works well for weighted graphs.

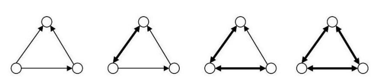

Fig. 4. Four types of cycles.

We are presently studying time windows on the graph of telephone calls in order to identify stable influent agents. The evolution of the graph depends on the period used to aggregate the data and the time shift that determines the next period. For example, one year of data can lead to 12 graphs aggregated over one month without overlap. In addition to evolving nodes and edges, the nodes can also have an evolving state. For instance, they may or may not have some technology.

This raises the question of sensitivity of social leaders. It appears that edges that participate to many cycles are stable, and thus most of the unstable edges belong to a few triangles. So they perturb hardly the SDs and eventually the SLs. The disappearance/appearance of some nodes may also affect considerably the social leaders. For that reason, we need suitable time windows that allow to find social leaders over long period of time.

References

1. Newman MEJ (2003) The structure and function of complex networks. SIAM Review 45, 167-256
2. Kamada T and Kawai S (1989) An Algorithm for Drawing General Undirected Graphs. Information Processing Letters 31, 7–15
3. Harel D, and Koren Y (2000) A fast multi-scale method for drawing large graphs. In Proc. 5th Working Conference on Advanced Visual Interfaces (AVI'00), A. Press, Ed., pp. 282–285. Appl Polym Sci 64:1079–1090
4. Clauset A, Newman MEJ, and Moore C (2004) Finding community structure in very large networks. Phys. Rev. E 70, 066111