How to exploit structural properties of dynamic networks to detect nodes with high temporal closeness

Marwan Ghanem[1], Clémence Magnien[1], Fabien Tarissan[2]

[1] Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, LIP6, F-75005 Paris, France

`firstname.lastname@lip6.fr`

[2] Université Paris-Saclay, CNRS, ENS Paris-Saclay, ISP UMR 7220, France

`fabien.tarissan@ens-paris-saclay.fr`

Abstract

The ability to detect important nodes in temporal networks has been investigated lately. This has been a challenge on both the theoretical aspects as well as computational ones. In this study we propose and evaluate different strategies to detect nodes that have high temporal closeness.

Keywords : *Temporal closeness, Sampling, Dynamic network*

1 Introduction and definitions

Evaluating the importance of nodes in complex networks has been an interesting question for a long time. Several measures of importance have been introduced, such as degree, closeness or betweenness centrality. As complex networks have grown in size, approximation methods have been introduced. One of the first method to approximate centrality was introduced in [6]. They consider k source nodes selected randomly, from which they compute the shortest-paths with all other nodes of the network. Since then, several methods have been proposed to help selecting the source nodes [4] or the target nodes [8] in order to reduce the computation required to estimate the closeness and betweenness centrality Those studies all consider a single and static network. However, most of real application involve networks whose structure evolves with time. This led the community to propose adaptation of centrality metrics to assess the importance of nodes through time [7, 11]. This temporal dimension makes the computation more demanding, making methods for approximating centrality metrics even more essential. In this study, we study how structural properties of dynamic networks can be exploited to detect nodes that have a high *temporal closeness centrality* [7].

More precisely, let $G = (V, E)$ be a dynamic network composed of a set V of nodes and a set E of temporal links of the form (u, v, t) where $u, v \in V$ and t is a timestamp. A temporal path from u to v starting at time t_s is given by a sequence of links $(u, v_0, t_0), (v_0, v_1, t_1), \ldots, (v_{k-1}, v, t_k)$ such that $t_0 > t_s$ and, for all $i, i = 0..k-1$, $t_i < t_{i+1}$.

Such a path is a *shortest path* if it has the least duration $(t_k - t_s)$ among all paths from u to v starting at time t_s. The *(temporal) distance* from u to v at time t_s is then the duration of such a shortest path (denoted $d_{t_s}(u,v)$)[1]. Following the classical definition of the closeness of a node in a static network, the *temporal closeness* of a node u at time t is defined by:

$$C_t(u) = \sum_{v \neq u} \frac{1}{d_t(u,v)}$$

It measures the importance of node u at time t in the dynamic network. In order to assess what nodes are important at time t, one can rank the nodes according to their temporal closeness at time t and consider for instance the top 25% rankings. This enables in turn to compute for each node u its total duration spent in the top 25% rankings (denoted by $Dur_{top}(u)$). It measures the *global* importance of node u in G. The purpose of the present study is to propose strategies to detect which nodes are globally important without relying on the exact computation of the temporal closeness of all nodes at all time instant.

2 Strategies and results

In order to detect globally important nodes, we propose to first compute global properties of the nodes that can easily be extracted either from the aggregated graph $G_A = (V, E_A)$ (with $E_A = \{(u,v) | \exists t, (u,v,t) \in E\}$) or from an analysis of the temporal activity. For every node u, we compute its closeness centrality $CC(u)$, its degree centrality $DC(u)$ and its number of links $NL(u)$ – all computed on G_A – as well as its duration of activity $DU(u)$[2] and its average inter-contact duration time $LD(u)$[3]. Then we propose:
Parameter based strategy (P_1/P_2): we consider the rankings given by mixing the importance measured by P_1 and P_2 defined by: $R(u) = \alpha \times \texttt{rank}(P_1(u)) + (1-\alpha) \times \texttt{rank}(P_2(u))$ with $\alpha \in [0:1]$[4] and where $\texttt{rank}(P)$ is the rank provided by property P.
Parameterless strategy (PS): we only take into account the number of links and the duration of activity: $R(u) = \texttt{rank}(NL(u) \times DU(u))$.
In order to assess the relevance of each strategy (and for any α), we compute the number of nodes correctly detected as important[5] in the top k nodes (for $k \in [1..n]$) and denote this vector as the *hit rate* vector. The hit rate vector of a perfect strategy would then be equal to $[1, 2, .., n]$. From these vectors we can compute the distance between any strategy and the perfect strategy and normalize it by the worse case strategy. Formally, we define the score of a strategy by: $score(S) = 1 - \frac{distance(perfect\_strategy, S)}{distance(perfect\_strategy, worse\_case)}$
Figure 1 shows the scores for all the strategies (with different values for α) when applied on nine datasets whose characteristics are provided in Table 1. We observe that in most cases NL/DU,DU/LD and PS score higher than other combinations as well as any pure static centralities. They are much closer to a perfect strategy or ground truth than any

[1] $d_{t_s}(u,v) = \infty$ if there is no path between u and v.
[2] the difference between that last and the first activity.
[3] the average time between two consecutive links involving u.
[4] note that $\alpha = 1$ implies that only P_1 is considered.
[5] we consider the exact computation of the temporal closeness as the ground truth.

Datasets	Type	#Nodes	#Edges	Duration	Ref
Enron	Email	151	47 088	3 years	[10]
Radoslaw	Email	168	82 876	9 months	[9]
DNC	Email	1891	39 264	2.6 years	[1]
HashTags	Social Network	3 048	100 429	22 days	-
Facebook	Social Network	8 977	66 153	1 year	[12]
Article Tags	Social Network	2902	571 877	10 years	-
Reality Mining	Movement	96	1 M	9 month	[5]
Taxi Rome	Movement	158	241 736	1 day	[3]
Primary	Movement	242	125 773	1.5 days	[2]

TAB. 1: Dataset, Type, Number of nodes, Number of links, Duration

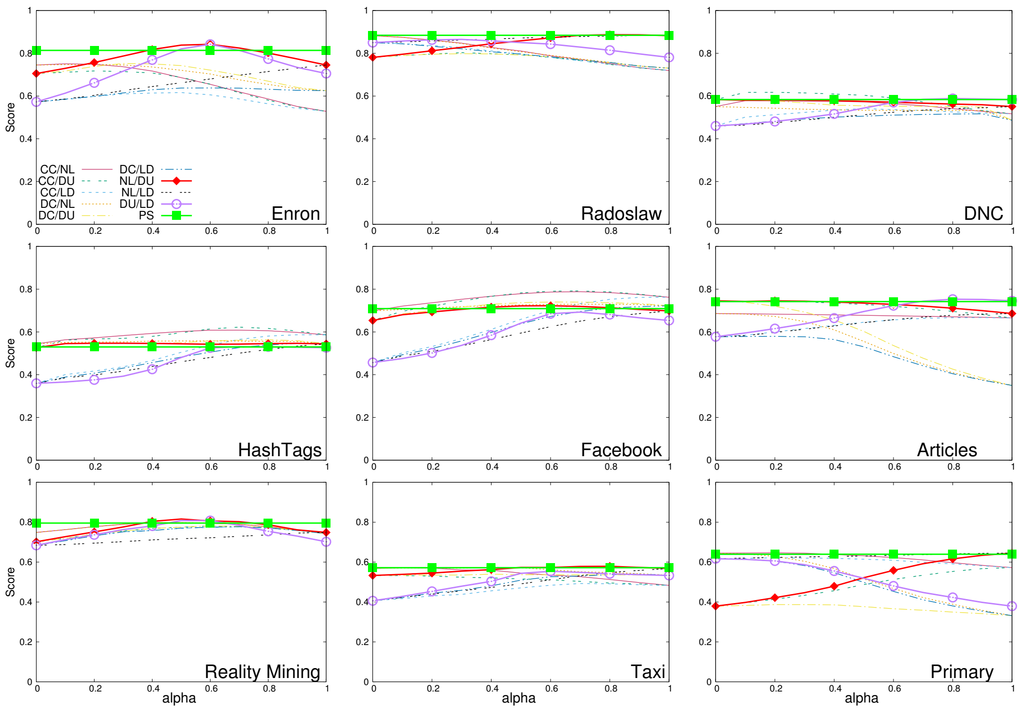

FIG. 1: Score for each strategy on the nine datasets

other strategies. In addition, we can observe that datasets of same nature lead to similar α value for the best strategies.

3 Conclusions

In this study we proposed different strategies that rely on global properties of nodes to detect nodes with high temporal closeness centrality. In most cases, three strategies present the best results. They all take into account temporal properties of the nodes.

This work is a first step to adapt recent technics [6, 4, 8] to approximate the importance of nodes in dynamics networks.

Acknowledgements

This work is funded in part by the European Commission H2020 FETPROACT 2016-2017 program under grant 732942 (ODYCCEUS), by the ANR (French National Agency of Research) under grants ANR-15-CE38-0001 (AlgoDiv) and ANR-13-CORD-0017-01 (CODDDE), by the French program "PIA - Usages, services et contenus innovants" under grant O18062-44430 (REQUEST), and by the Ile-de-France program FUI21 under grant 16010629 (iTRAC).

References

[1] Dnc co-recipient network dataset – KONECT, September 2016.

[2] A.L. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[3] Lorenzo Bracciale, Marco Bonola, Pierpaolo Loreti, Giuseppe Bianchi, Raul Amici, and Antonello Rabuffi. CRAWDAD dataset roma/taxi (v. 2014-07-17). Downloaded from `https://crawdad.org/roma/taxi/20140717`, July 2014.

[4] Ulrik Brandes and Christian Pich. Centrality estimation in large networks. *International Journal of Bifurcation and Chaos*, 17(7):2303–2318, 2007.

[5] Nathan Eagle and Alex (Sandy) Pentland. CRAWDAD dataset mit/reality (v. 2005-07-01). Downloaded from `https://crawdad.org/mit/reality/20050701`, July 2005.

[6] David Eppstein and Joseph Wang. Fast approximation of centrality. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '01, pages 228–229, Philadelphia, PA, USA, 2001. Society for Industrial and Applied Mathematics.

[7] Clémence Magnien and Fabien Tarissan. Time evolution of the importance of nodes in dynamic networks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ASONAM '15, pages 1200–1207, New York, NY, USA, 2015. ACM.

[8] Arun S. Maiya and Tanya Y. Berger-Wolf. Online sampling of high centrality individuals in social networks. In *Proceedings of the 14th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining - Volume Part I*, PAKDD'10, pages 91–98. Springer-Verlag, 2010.

[9] Radosław Michalski, Sebastian Palus, and Przemysław Kazienko. Matching organizational structure and social network extracted from email communication. In *Lecture Notes in Business Information Processing*, volume 87, pages 197–206. Springer Berlin Heidelberg, 2011.

[10] Jitesh Shetty and Jafar Adibi. Discovering important nodes through graph entropy the case of Enron email database. In *Proceedings of the 3rd international workshop on Link discovery - LinkKDD '05*, pages 74–81, New York, New York, USA, August 2005. ACM Press.

[11] Taro Takaguchi, Yosuke Yano, and Yuichi Yoshida. Coverage centralities for temporal networks. *The European Physical Journal B*, 89(2):35, 2016.

[12] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi. On the evolution of user interaction in Facebook. In *Proc. Workshop on Online Social Networks*, pages 37–42, 2009.