How to exploit structural properties of dynamic networks to detect nodes with high temporal closeness
Marwan Ghanem, Clémence Magnien, Fabien Tarissan

To cite this version:
Marwan Ghanem, Clémence Magnien, Fabien Tarissan. How to exploit structural properties of dynamic networks to detect nodes with high temporal closeness. Cologne-Twente Workshop on Graphs and Combinatorial Optimization 2018 (CTW’18), 2018, Paris, France. hal-01915209

HAL Id: hal-01915209
https://hal.science/hal-01915209
Submitted on 7 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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, Approximation, 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 \(d_s(u, v)\)). 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 \(E_A = \{(u, v) \mid \exists t, (u, v, t) \in E\}\)) 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 \(DU(u)\) and its average inter-contact duration time \(LD(u)\). 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:

\[
R(u) = \alpha \times rank(P_1(u)) + (1 - \alpha) \times rank(P_2(u))
\]

with \(\alpha \in [0:1]\) and where \(rank(P)\) is the rank provided by property \(P\).

**Parameterless strategy** \((PS)\): we only take into account the number of links and the duration of activity:

\[
R(u) = rank(NL(u) \times DU(u)).
\]

In order to assess the relevance of each strategy (and for any \(\alpha\)), we compute the number of nodes correctly detected as important\(^4\) 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:

\[
\text{score}(S) = 1 - \frac{\text{distance}(\text{perfect\_strategy}, S)}{\text{distance}(\text{perfect\_strategy}, \text{worse\_case})}
\]

Figure\(^5\) shows the scores for all the strategies (with different values for \(\alpha\)) 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. \(d_s(u, v) = \infty\) 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.
<table>
<thead>
<tr>
<th>Datasets</th>
<th>Type</th>
<th>#Nodes</th>
<th>#Edges</th>
<th>Duration</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enron</td>
<td>Email</td>
<td>151</td>
<td>47 088</td>
<td>3 years</td>
<td>[10]</td>
</tr>
<tr>
<td>Radoslaw</td>
<td>Email</td>
<td>168</td>
<td>82 876</td>
<td>9 months</td>
<td>[9]</td>
</tr>
<tr>
<td>DNC</td>
<td>Email</td>
<td>1891</td>
<td>39 264</td>
<td>2.6 years</td>
<td>[1]</td>
</tr>
<tr>
<td>HashTags</td>
<td>Social Network</td>
<td>3 048</td>
<td>100 429</td>
<td>22 days</td>
<td>-</td>
</tr>
<tr>
<td>Facebook</td>
<td>Social Network</td>
<td>8 977</td>
<td>66 153</td>
<td>1 year</td>
<td>[12]</td>
</tr>
<tr>
<td>Article Tags</td>
<td>Social Network</td>
<td>2 902</td>
<td>571 877</td>
<td>10 years</td>
<td>-</td>
</tr>
<tr>
<td>Reality Mining</td>
<td>Movement</td>
<td>96</td>
<td>1 M</td>
<td>9 month</td>
<td>[5]</td>
</tr>
<tr>
<td>Taxi Rome</td>
<td>Movement</td>
<td>158</td>
<td>241 736</td>
<td>1 day</td>
<td>[3]</td>
</tr>
<tr>
<td>Primary</td>
<td>Movement</td>
<td>242</td>
<td>125 773</td>
<td>1.5 days</td>
<td>[2]</td>
</tr>
</tbody>
</table>

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

![Graphs](image_url)

**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 \( \alpha \) value for the best strategies.

### 3 Conclusions

In this study we proposed different strategies that rely on global characteristics 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 (ODYCEUS), 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