Time Evolution of the Importance of Nodes in dynamic Networks

Clémence Magnien and Fabien Tarissan
Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu 75005 Paris, France
Phone: +33(1)44278792 / Fax: +33(1)44278810
Email: first.lastname@lip6.fr

***Abstract*—For a long time now, researchers have worked on defining different metrics able to characterize the importance of nodes in networks. Among them, *centrality* measures have proved to be pertinent as they relate the position of a node in the structure to its ability to diffuse an information efficiently. The case of dynamic networks, in which nodes and links appear and disappear over time, led the community to propose extensions of those classical measures. Yet, they do not investigate the fact that the network structure evolves and that node importance may evolve accordingly. In the present paper, we propose temporal extensions of notions of centrality, which take into account the paths existing *at any given time*, in order to study the time evolution of nodes' importance in dynamic networks. We apply this to two datasets and show that the importance of nodes does indeed vary greatly with time. We also show that in some cases it might be meaningless to try to identify nodes that are consistently important over time, thus strengthening the interest of temporal extensions of centrality measures.**

Keywords*—centrality, network dynamics, temporal paths, node importance

I. Introduction

Scientists studying complex networks have been interested for a long time in the question of evaluating the importance of a node. This has lead to the introduction of several measures of importance, such as for instance degree, closeness or betweenness centrality, Katz centrality, or PageRank.

Most centrality measures are based on the study of paths in the network: a node will be important for instance if the paths from it to other nodes are short, or if it lies on shortest paths between many pairs of nodes. One motivation for this is that links can act as a dissemination medium for some phenomena occurring on the network. For instance individuals can exchange information when they communicate, or a message can be forwarded from computer to computer until it reaches its destination.

Researchers have acknowledged for some time that networks are dynamic in nature: nodes and links come and go with time. This has led to a stream of works aiming at understanding and modelling these dynamics. In particular in the case of centrality, some works have been concerned with efficiently updating the centrality values of the nodes when a change occurs in the network. In many cases however, the time scale at which the network evolves is the same as the one at which a dissemination phenomena may occur on the network. This is the case for instance when a disease propagates among individuals when they are in contact, or when an information is disseminated by email messages.

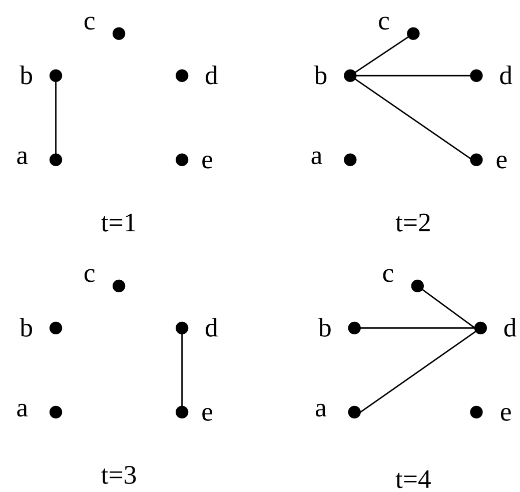

Fig. 1. A small example of a dynamic network. The links existing at time $t = 1$ are shown on the top left corner, the ones existing at $t = 2$ in the top right corner, and so on.

In this case, it becomes necessary to consider paths that are not instantaneous but instead are spread between a beginning and an ending time, while respecting the network dynamics. Such a *temporal* path follows links that happen one after the other (see Section III for a rigorous definition).

In this case, our key argument is that paths change during the network time span, and thus the importance of nodes varies. Consider indeed the toy example of Figure 1. This small network composed of five nodes evolves during four distinct time steps. One can see that, intuitively, the importance of node b is much stronger at time $t = 1$ than at time $t = 3$. Indeed, at time $t = 1$ it forms a bridge between node a and nodes c, d and e, thanks to the links that exist at time $t = 2$.

Several works have introduced extensions of centrality notions for the case of dynamic networks. However, most of these works consider only paths starting at the beginning of the dataset; they obain in this way a single figure for node centrality, representative only of what happens at very early times. Other works consider paths throughout the dataset time span, but still consider that node importance can be represented by a single figure rather than by a time-evolving metric.

Though it is quite straightforward to extend these metrics in order to consider the time evolution of node centrality, no work up to our knowledge has attempted to study this question.

In this paper, we study a natural extension of the closeness centrality to the case where paths may start at any time during the network's time span. This temporal closeness characterizes the importance of any node *at any given time*. We study two datasets and observe that:

1) node importance *does* vary with time, and therefore capturing the global importance of a node with an aggregate value might be misleading;
2) in some cases the dynamic of the network is such that it is meaningless to identify nodes more important than others; in other cases, temporal closeness may help identifying a node consistently important for the whole network time span;

This work is organized as follows. First we present the existing work related to the notion of centrality in static and dynamic networks (Section II) before providing the definitions necessary to the present study (Section III). Then we present the two datasets (Section IV) on which we apply the proposed metrics and present the obtained results (Section V). Finally, we conclude the paper with some perspectives (Section VI).

II. Related work

Many papers have studied the importance of nodes in *static* networks, i.e. networks that don't evolve with time. Among the metrics that have been introduced, one may cite the degree centrality, closeness centrality [1], betweenness centrality [2] and the Katz centrality [3]. Closeness and betweenness centrality are based on the shortest paths, while the Katz centrality takes into account the paths of all lengths between two nodes.

Some papers who have studied dynamic networks have been concerned with efficiently computing the static centrality at all times. For instance, Kas *et al.* [4] propose an algorithm that, given the knowledge of the distances between all pairs of nodes and given a network change (edge appearance of disappearance), computes the new distance values (which allows the computation of distance-related centrality measures) by updating the previous values rather than computing them all from scratch again. This is relevant, e.g. in contexts where the network evolves at a much slower scale than the one on which disseminations take place.

However, in many contexts the dissemination phenomena in the network happen on the same time scale as the network evolution. It then becomes necessary to consider *temporal* paths [5], [6], i.e. link sequences that are time-respecting, as opposed to paths composed of links that all exist at the same time. For instance, in the dynamic network of Figure 1, there is a temporal path from node a to node e going through the link (a, b) at $t = 1$ and the link (b, e) at time $t = 2$.

Several definitions of temporal paths have been studied in the literature. Some of them can be computed more easily than others. Whitbeck *et al.* [7] propose an efficient algorithm to approximate the existence of paths in the most difficult case, and show that the study of the notion of reachability, i.e. which nodes can be reached from which ones, and at which times in the network's time span, provides enlightening insight on the network's dynamics.

Notions of centrality taking into account temporal paths have also been introduced.

Nicosia *et al.* [8] introduce notions of temporal closeness and betweenness centralities. Their definition of a shortest path however considers only paths whose starting point is at the beginning of a dataset's time span.

Some propositions acknowledge that the distances between nodes, and therefore nodes' importance, vary with time [9], [10], [11], [6] . However, in practice they still represent the varying importance of a node by a single value that is supposed to represent its overall importance throughout the network global time span.

Several papers introduce and study a variant of the Katz centrality [12], [13]. Among those, Lerman *et al.* [14] acknowledge the fact that node importance may evolve with time, but no systematic study is performed. Moreover, the introduced metric is dependant on parameters defining what are considered as relevant path lengths and path durations, which complicates the analysis.

Finally, Cotsa *et al.* [15] notice that not all time instants are equivalent in a dynamic network, and introduce the notion of *time centrality*; this is a measure of how fast a dissemination process can reach a significant portion of the nodes at a given time t. However, this notion does not analyze the importance of individual nodes in the dissemination process, which is our goal in this paper.

All in all, and to the best of our knowledge, if many papers acknowledge the fact that the temporal evolution of networks impact the value of centrality measures and propose variations of standard metrics to account for the dynamics, no paper propose a complete and systematic study of the evolution of centrality measures for all nodes and all time steps of the network's evolution. Instead, they all propose average computations either over all nodes of the networks, either over all the network time span. We argue in this paper that discarding either of those two aspects lead to severe misunderstanding of the real nature of node's importance in the context of dynamic networks

III. Definitions

A dynamic network $G = (V, E)$ consists of a set V of nodes [1] and a set E of timed links of the form (u, v, t) where $u, v \in V$ and t is a timestamp. Throughout the paper we consider networks as undirected, i.e. a link (u, v, t) is equivalent to a link (v, u, t).

A temporal path in a dynamic network consists of:

- a starting time t_s, and
- a sequence of links $(u_0, v_0, t_0), (u_1, v_1, t_1), \ldots, (u_k, v_k, t_k)$

such that:

1) for all $i, i = 0..k-1$, $u_{i+1} = v_i$;
2) for all $i, i = 0..k-1$, $t_i < t_{i+1}$.
3) $t_0 \geq t_s$;

We say that such a path is a path from u_0 to v_k starting at time t_s. Its *duration* is equal to $t_k - t_s$. We will say that

[1] We assume that the set of nodes does not evolve with time.

a path from u to v starting at time t_s is a shortest path if it has the least duration among all paths from u to v starting at time t_s. We define the distance from u to v at time t to be the duration of a shortest path from u to v starting at time t, and we denote it by $d_t(u, v)$. If there is no path from u to v starting at time t, we consider that $d_t(u, v) = \infty$.

Note that a path starting at time t might imply waiting times at all nodes, including the first one, in the same way that a person starting at a given time a train trip with connections must wait for the train in the first station, and then at each connecting station.

For example, in the dynamic network of Figure 1, there are two temporal paths from e to b starting at time $t = 2$. The first one consists of the single link $(e, b, 2)$, and its duration is 0; the second one consists of the links $(e, d, 3)$ and $(d, b, 4)$ and its duration is 2. The temporal distance from e to b at time 2 is therefore $d_2(e, b) = 0$. Note that the temporal distance from e to b at time 1 is $d_1(e, b) = 1$: the temporal shortest path has the same link sequence than the one starting at $t = 2$ (it is the single link $(e, b, 2)$), but the starting time is different. Since there is no temporal path from c to e starting at time $t = 3$, $d_3(c, e) = \infty$.

Several variants have been introduced in the literature, most notably concerning the constraint $t_i < t_{i+1}$. Some variants weaken it to $t_i \leq t_{i+1}$, while others strengthen it to $t_i \leq t_{i+1} + \delta$, where δ is a parameter representing the time needed to send a message along a link. The relevance of these variants depends on the context. See [7] for more details. Preliminary work shows that this has little influence on the results.

It is worth noticing that our definition of a dynamic network consists of links without a duration. Again, this is relevant in some contexts (e.g., an email is sent at a precise time) and less in others (a phone call has an intrinsic duration). Notice however that our notion of a temporal path can be easily adapted in the latter case; the condition becomes that a path going through node u at time t_i can continue to a node v provided that there is a link between u and v that ends after t_i.

We recall that the closeness of a node u in a non-evolving network is defined as [1]:

$$\sum_{v \neq u} \frac{1}{d(u, v)},$$

where $d(u, v)$ is the classical graph distance.

The average of the closeness of all nodes has been defined as network *efficiency* [16] [2].

Though some extensions of the closeness have been defined for the case of dynamic networks [11], [8], their goal is not to take fully into account the fact that temporal distances vary according to the paths' starting time. We therefore here define the temporal closeness of a node u at time t as:

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

and we define the temporal efficiency $E_t(G)$ of network G as the average over all nodes of the temporal closeness at time t.

Building upon this notion of efficiency, we can study another metric that quantifies the impact a node has on a network. The notion of *delta-centrality* [17] characterizes how much a given node (or group of nodes) impacts the efficiency. It is defined as the relative change of the network efficiency when the considered node (or group of node) is removed from the network.

Following this, we study the extension of the delta-centrality to the dynamic case. The *temporal delta-centrality* of a node v at time t is defined by:

$$\frac{E_t(G) - E_t(G \backslash v)}{E_t(G)},$$

where $G \backslash v$ is the network obtained from G by removing node v and all its adjacent links.

The program we used to compute the above metrics is available [18].

IV. DATA SETS

In order to study the behaviors of the metrics introduced above, we study two datasets that present different characteristics and come from two very different contexts:

- Rollernet [19]: this dataset was collected during a rollerblade tour in Paris in August 2006. The tour is a weekly event and gathers approximately 2500 participants. Among these, 62 were equipped with wireless sensors recording when they are at a communication distance from one another. The dataset therefore contains the proximity links between the persons carrying the sensors. The total dataset duration is approximately 2 hours and 45 minutes (with a break of approximately 30 minutes).
- Enron [20]: this dataset contains the 252 759 emails that 151 Enron employees exchanged during three years. It records information on the senders, receivers, and the moment they were sent. Note that by nature, the links are directed but for a fair comparison with the Rollernet dataset, we treated them as undirected in the present study.

Before studying the importance of nodes (see next section), it is enlightening to make some global observations related to the dynamic of the networks.

To do so, we present in Figure 2 the fraction of pairs of nodes for which there is a temporal path starting at the beginning of the dataset and ending before time t, as a function of t. In other words, it represents the proportion of pairs of nodes that are reachable from one another at time t or sooner. Notice that, though the shapes of the plots are similar, the time scales are very different: for Rollernet, all pairs of nodes that are eventually connected are so within less than half an hour, i.e. quite early in the dataset. For Enron, on the other hand, the time scale is larger and we can see that less than 10% of the pairs of nodes are reachable from one another after one year; new pairs become reachable for the whole dataset duration, and some paths only end very close to the end of the dataset.

[2] There are several variants of the closeness in the litterature, depending on a normalisation constant, and the efficiency depends of this. Since for both metrics this is a constant that does not vary with time, we use the simplest version for the closeness which is the sum of the inverses of the distances.

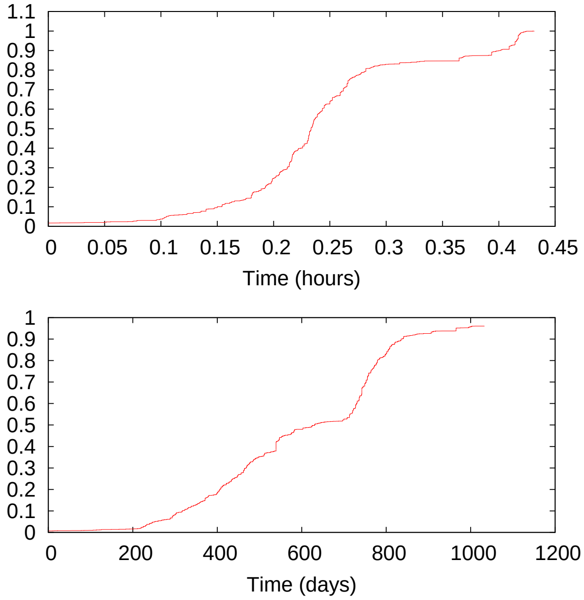

Fig. 2. Fraction of pairs of nodes for which there exists a temporal path starting at the beginning of the dataset and ending at time t or sooner, as a function of t. Top: Rollernet; Bottom: Enron.

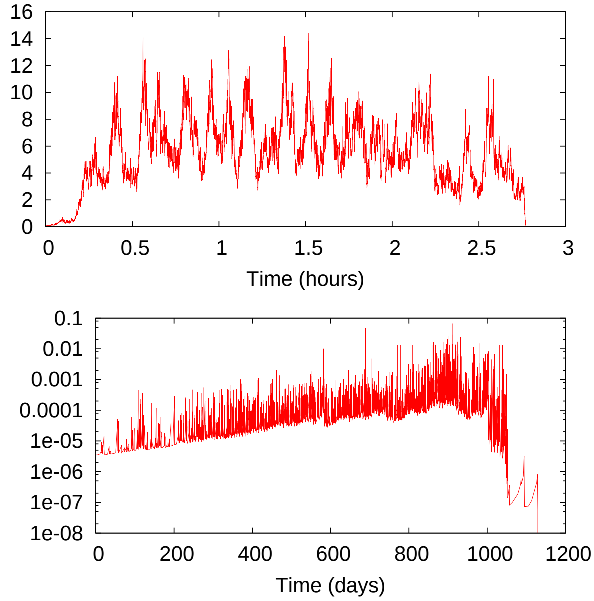

Fig. 3. Time-evolution of the temporal efficiency. Top: Rollernet; Bottom: Enron.

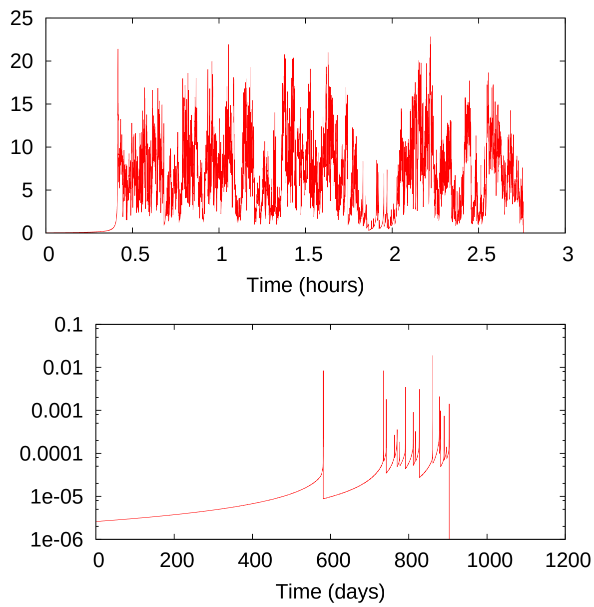

Fig. 4. Time-evolution of the temporal closeness of randomly chosen nodes in our datasets. Top: Rollernet; Bottom: Enron.

This difference between the datasets does not come from an artifact at the beginning of Rollernet: we have observed that we can find short paths from any node to any other node at almost any starting time (except close to the end). This seems to indicate that nodes will be overall less important in Rollernet than in Enron. Indeed, in the first case we may expect that a given node's closeness will be only marginally larger than an other one's, whereas in Enron we may expect that *all* paths to a given node must go through the same node, which will then be quite important. Our observations in the next section confirm this intuition.

V. Results

We now present the results obtained using the different notions introduced in Section III on the two dynamic networks presented above.

A. Temporal efficiency and temporal closeness over time

Figure 3 presents the time evolution of the temporal efficiency for each dataset. We can see that the value fluctuates widely for both of them. Notice that for Enron, even though it fluctuates, the efficiency tends to increase with time [3]. This is caused by the fact that, as shown previously, the number of pairs that are reachable from one another increases with time throughout the dataset duration. The final collapse stems from the fact that, towards the end of the dataset, less and less temporal paths exist.

We expect that these fluctuations in the efficiency will impact the individual nodes' closeness.

Turning to the temporal closeness, we first present in Figure 4 the time evolution of this metric for randomly chosen nodes in our datasets. It appears clearly that the closeness of a node can be very bursty. Values vary on a wide range in both cases, and along several orders of magnitude in the case of Enron. This burstiness makes the plots difficult to interpret, and highlights our claim that the importance of a node varies greatly with time.

However, the fact that a node's closeness fluctuates with time does not mean that this node does not have a relatively large (or small) closeness overall. To evaluate this, we need to compare the values between nodes. We do this in the next section.

[3]Note that we used a log scale on the y-axis for this plot. The peaks spanning several orders of magnitude make it unreadable with a linear scale.

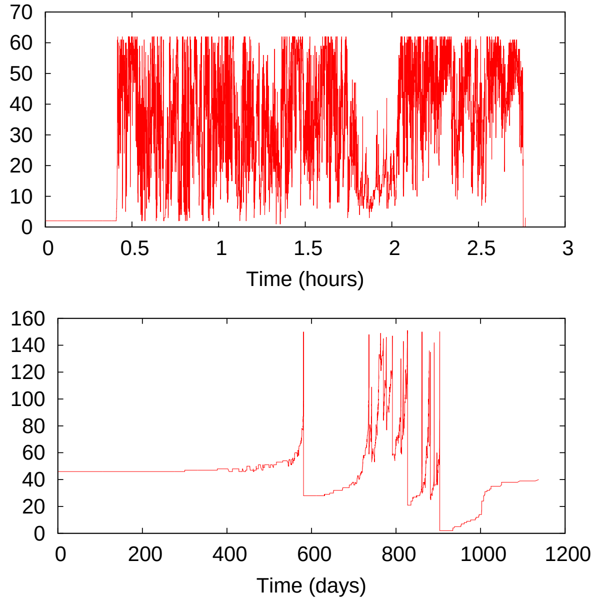

Fig. 5. Time-evolution of the rank of randomly chosen nodes in our datasets. Top: Rollernet; Bottom: Enron.

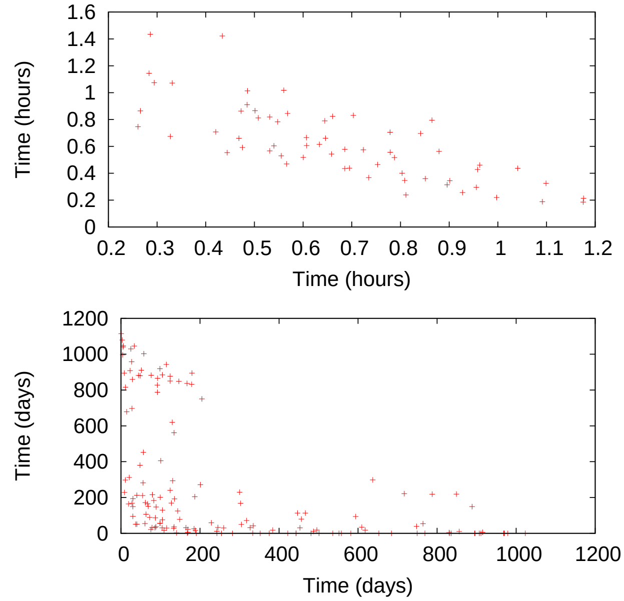

Fig. 6. Total duration for which rank is in the 25% lowest (x-axis) vs total duration for which rank is in the 25% highest (y-axis). Top: Rollernet; Bottom: Enron.

B. On the relative importance of a node

In order to get a better sense of whether a node consistently has a large closeness with respect to other nodes, we proceeded in the following way: for each time step t, we sorted all nodes by increasing order of their temporal closeness at time t. This gives to each node a rank that varies with time: node x has rank 1 at time t if it has the lowest temporal closeness among all nodes at time t, and rank n (where n is the number of nodes) if it has the highest temporal closeness.

Figure 5 shows the time evolution of the ranks of the same randomly chosen nodes as Figure 4. We can observe several things. First, as expected, there seems to be a correlation between closeness and rank: at many times, the higher the closeness, the higher the rank. However, this correlation is not perfect: for the Enron node, for instance, there is a period of increasing closeness between two peaks after t=800 days. The closeness at this time is larger than at times 0-600 days, but the corresponding rank is lower. We also observe an artefact for large times in the Enron case: we observe that the closeness drops to 0 after approximately 900 days; however the rank increases during this period. This is due to the finite duration of the dataset. Since fewer and fewer temporal paths exist between pairs of nodes, the closeness of more and more nodes drops to zero (we observed the corresponding decrease in the efficiency in Figure 3). The rank among those nodes is then arbitrary.

Finally, we observe different behaviors between the Rollernet and the Enron nodes. While the Rollernet node's rank fluctuates between very low and high ranks (except at the beginning of the trace), the Enron node's rank is rather stable for a significant part of the dataset's duration, before starting fluctuating (but not as drastically as the Rollernet one). In conclusion, neither of these nodes has a high or low closeness globally. The Enron node has a low closeness for a significant duration of the dataset, but still reaches very high ranks at some times, at which it plays an important role in the network connectivity. The Rollernet node has no clear period within which its rank is rather stable.

C. On the global importance of some nodes

In order to study all nodes in a more systematic manner, we computed, for each node in each dataset, the total duration during which its rank was among the 25% lowest, and the total duration during which its rank was among the 25% highest. We plot in Figure 6 these two quantities for each node for the Rollernet (top) and Enron (bottom) case. Notice that the shape of the scatterplot is limited by the fact that the sum of both coordinates cannot exceed the dataset duration.

In both datasets we observe that all nodes are not equivalent: most nodes are in the top 25% ranks for a longer duration than they are in the bottom 25%, or conversely. However, this does not happen with the same magnitude in both datasets. In Rollernet, the maximum total duration for which a node is in the top or bottom 25% ranks is approximately 1.5 hours, which represents approximately half the dataset duration. This means that no node has a globally high or low closeness throughout the whole dataset duration. In Enron however, some nodes are predominantly either in the highest or lowest 25% ranks during almost all the dataset's time span, meaning that these nodes consistently have a high (or low) temporal closeness *comparatively* to all other nodes.

In order to deepen the study of whether some nodes play an overall important role in the network (or conversely, have consistently a low impact), we identfied the node for which the rank was in the top (resp. bot) 25% for the longest period of time, for each dataset. We plot the time evolution of these nodes' rank in Figures 7 (Rollernet) and 8 (Enron).

Again, we observe different behaviors between the datasets. For Rollernet, the longest duration a node is in the 25%

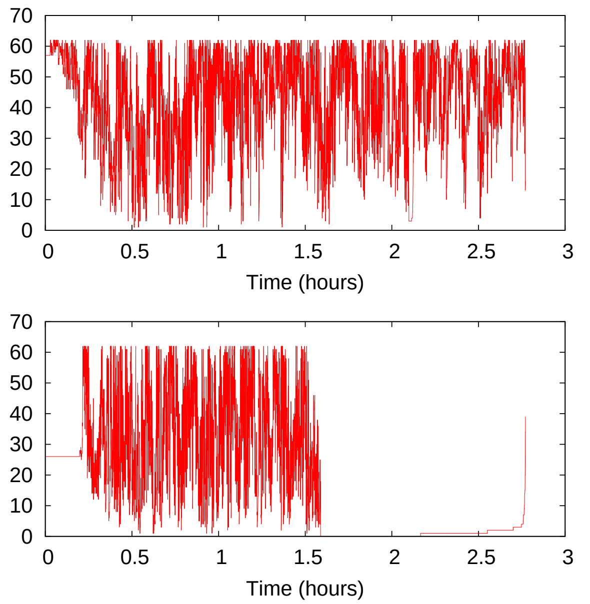

Fig. 7. Time evolution of the rank in the Rollernet dataset, for the node with the longest duration in the highest 25% ranks (top) and the node with the longest duration in the lowest 25% ranks (bottom).

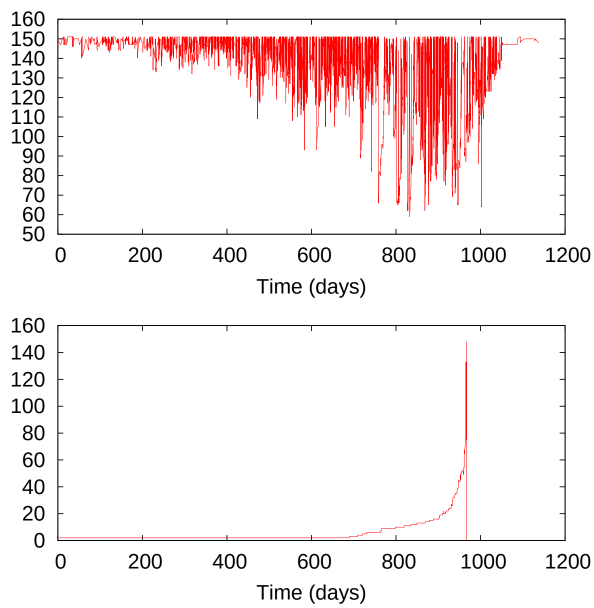

Fig. 8. Time evolution of the rank in the Enron dataset, for the node with the longest duration in the highest 25% ranks (top) and the node with the longest duration in the lowest 25% ranks (bottom).

highest ranks is approximately 1 hour and 10 minutes, which is significantly less than half the duration of the dataset. We can moreover observe that the corresponding node's rank fluctuates a lot (Figure 7 top) and never stays in the highest ranks during a significant amount of time.

The case of the node with the most time in the lowest ranks is different (Figure 7 bottom). It has a very low rank in approximately the second half of the dataset. Indeed, this node is only active for the first half of the dataset, and does not have any link in the second half. Its temporal closeness therefore drops to zero (but, as seen in Figure 4 it has then an arbitrary rank among all nodes with a null closeness). However, it is quite interesting to observe that, while this node is active, its rank fluctuates a lot and often reaches very high ranks, meaning that it is among the most influential nodes at these times. In fact, this node spends slightly more than 15 minutes in total among the 25% highest ranks, while being active only 1h30 in total. Altogether, the conclusion for Rollernet is that no node is globally important or unimportant for a long time; global importance may not be a relevant notion for this dataset.

By contrast, the situation is quite different for the Enron case. Although the rank of the node with the longest duration in the highest ranks fluctuates with time (Figure 8, top), we can observe that it consistently stays within the highest 50% ranks for approximately the first half of the dataset's time span. Moreover, even though its rank tends to decrease in the second half, it never goes in the lowest 33% ranks. This node is therefore globally important in the dataset.

In the same way, the node with the longest duration in the lowest ranks (Figure 8, bottom) is consistently unimportant. It is quite interesting to notice that there is a point at which this node reaches a very high rank (rank 149 out of 151 nodes). A manual study indicates that this node is actually inactive for most of the dataset time span. Its only activity consists in exchanging two messages with the same node, shortly before 1 000 days. It is striking that these two links are enough to bring this node to the third highest rank among all nodes. This indicates that these links are very important in the network, and/or that the network dynamics undergoes a particular event at this time. We leave the detailed investigation of such phenomena to future work.

D. *Delta-centrality*

In order to further study the importance of a node, we turn now to the temporal delta-centrality as defined in section III, which quantifies the impact of a node on the network.

We made the same systematic study for delta-centrality as the one we did for closeness: we computed the delta-centrality for all nodes at each time step; then for each time-step we sorted the nodes by increasing order of the delta-centrality, thus ranking them from 1 (lowest temporal delta-centrality) to n (where n is the number of nodes, for the highest delta-centrality).

Doing so, we observed that the temporal closeness is a very good estimator for the delta-centrality [4]. In general, the higher the temporal closeness, the higher the temporal delta-centrality; the rankings and their time evolution are also very close for most nodes. In the same way, the duration a node's rank is amont the highest (resp. lowest) 25% ranks for the temporal closeness is very highly correlated with the duration for which its ranks is in the same range for the temporal delta-centrality. Figure 9 illustrates this. We observe that the correlations are

[4] Note that the delta-centrality computation is much more costly than the closeness centrality, as the delta-centrality computation requires the computation of all distances between all pairs of nodes, both in the original network and in the network where the considered node has been removed; the temporal closeness requires only to compute the distance from one node to all other nodes.

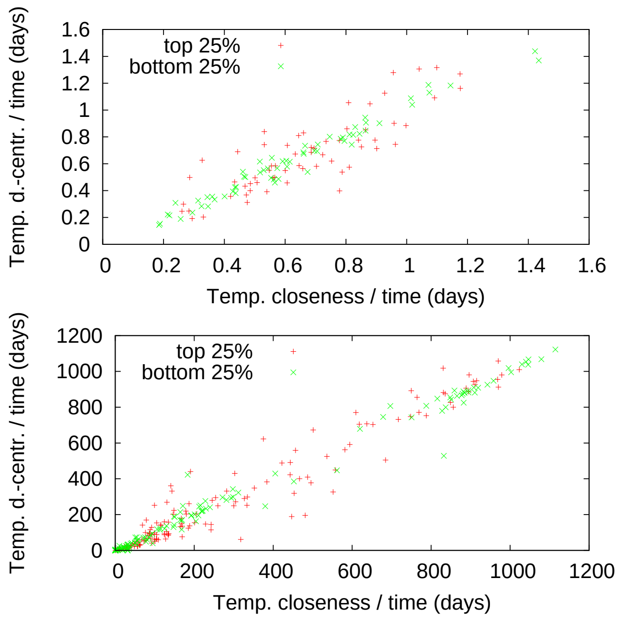

Fig. 9. Correlation between duration spent in lowest 25% or highest 25% ranks for temporal closeness (x-axis) and temporal delta-centrality (y-axis). Top: Rollernet; Bottom: Enron.

strong in both cases, and even stronger in the case of the lowest ranks.

We performed a detailed comparison of the temporal closeness and the temporal delta-centrality for individual nodes, and in particular for all nodes for which the correlation was the weakest. In such cases we observed in general that the corresponding nodes are not globally important or unimportant nodes: their coordinates in the scatteplots of Figure 6 are not very large. Moreover the time evolution of the rank for both centrality metrics are very similar in most cases, even though the ranks are not exactly the same. In most cases, the differences come from the fact that the rank is close to the 25% limit for a significant duration.

For a very small number of nodes, however, we observed a somewhat significant difference; in particular a few nodes tend to have a relative temporal delta-centrality that is higher than their relative temporal closeness. Such nodes therefore have a high impact on paths (when they are removed, the network temporal efficiency decreases), without being particularly close to other nodes.

These nodes therefore play a peculiar role in the network, and do not follow the general network behavior. We leave a detailed study of the causes of this phenomenon, as well as the identification of such anomalous nodes, to future work.

VI. Conclusion

Our central point in this paper is that temporal distances in dynamic networks may vary a lot with time; all distance-based centrality measures should therefore be considered as time-dependant, contrary to what has been mostly done in the litterature.

In order to investigate this, we studied the notion of temporal path *with a starting time*, and the corresponding notion of distance between two nodes *at a given time*. We then proposed temporal extensions of two importance measures, the closeness centrality and the delta-centrality. Using these notions, we studied two datasets coming from different contexts in order to investigate whether different network properties impact the observations.

Our observations can be summarized as follows:

1) node importance varies with time: a given node may be very important at one time, and not so important at another time; therefore it is not relevant to consider only aggregate values that summarize the importance of nodes on the whole network time span;
2) different datasets have different properties regarding node importance; for one of our datsets, the importance of all nodes fluctuates extremely rapidly between high and low values; it is meaningless in this case to state that one node is more important than another, except for a very limited time span; for our other dataset however, we find that some nodes are consistently important (or unimportant) for the whole network time span;
3) our studies have highlighted some specific nodes with atypical behaviors; this suggests that temporal centrality metrics could lead to methods for event and/or anomalous behavior detection.

Our work opens several interesting perspectives.

First, if the choice of two dataset stemming from very different contexts strengthen the conclusions drawn from the present study, we would like to apply the approach on more dataset involving dynamic networks. This would allow to confirm our findings and might help identify specific patterns of the evolution of node's importance that are context-dependant.

Following up on the precedent point, it would be very interesting for many real applications to be able to detect specific patterns in the evolution of centrality measures. This would indeed allow to predict which nodes are likely to be important in the future, which turns out to be of key importance for several applications, ranging from protocols of communication to recommendation systems.

On a more formal perspective, our definition of closeness of a node v relies on the computation of the distances from v to all other nodes. This is particularly relevant in our context, where we are concerned by the importance of a node in the dissemination process: a node will be important if it can reach many other nodes quickly. In other contexts however, the importance of a node v may be more closely related to the fact that the distances *from all other nodes to* v are short. This may be the case for instance in Web graphs, in which the importance of a page comes from the links towards it, not from its outgoing links. Comparing these two notions of closeness would lead to interesting insights.

We have seen that some of our observations allow to detect nodes that have an atypical behaviour, and/or moments where something unusual happens in the network's dynamics. We have observed this both when studying the time-evolution of the closeness of individual nodes, and when comparing different importance measures. This suggests that temporal centrality measures are relevant metrics when trying to detect anomalies in the network, which is a crucial question in many contexts [21]. In particular, it seems that nodes that are important with respect to one metric but not to another have a particularly interesting behavior. A systematic comparison of different metrics would therefore certainly lead to very interesting insight about the considered dataset.

In the same way, we have seen that some links play an extremely important role. It would be quite interesting to define a reliable method for identifying important links. On the one hand, this would lead to another approach for event detection in the network, complementary to the one sketched above. On the other hand, notions of link centrality have been successfully used in the case of static networks for community detection [22]. Using a notion of importance in a dynamic network for dynamic community detection is therefore a promising idea.

Finally, there is yet no consensus on relevant generative models for dynamic networks. Since we have observed that different networks have different properties regarding the temporal closeness centrality, this is probably an important ingredient to take into account when proposing a new model.

ACKNOWLEDGEMENTS

The authors would like to thank Thibault Rieutord for enlightening preliminary discussions on this work and in particular the algorithmic aspects.

This project was partly funded by a grant from the French program "PIA – Usages, services et contenus innovants" under grant number O18062-44430, and by the French National Research Agency contract CODDDE ANR-13-CORD-0017-01.

REFERENCES

[1] A. Bavelas, "Communication patterns in task-oriented groups," *Journal of the Acoustical Society of America*, vol. 22, p. 725730, 1950.

[2] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, pp. 35–41, 1977.

[3] L. Katz, "A new status index derived from sociometric index," *Psychometrika*, vol. 18, pp. 39–43, 1953.

[4] M. Kas, K. M. Carley, and L. Carley, "Incremental closeness centrality for dynamically changing social networks," in *IEEE/ACM International Conference on Advances in Social Netowrks Analysis and Mining (ASONAM)*. IEEE, 1013, pp. 1250–1258. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6785863

[5] B.-M. Bui-Xuan, A. Ferreira, and A. Jarry, "Computing shortest, fastest, and foremost journeys in dynamic networks," *International Journal of Foundations of Computer Science*, vol. 14, no. 2, pp. 267–285, Nov. 2003. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0129054103001728

[6] R. K. Pan and J. Saramäki, "Path lengths, correlations, and centrality in temporal networks," *Physical Review E*, vol. 84, no. 1, p. 016105, Jul. 2011. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.84.016105

[7] J. Whitbeck, M. D. de Amorim, V. Conan, and J.-L. Guillaume, "Temporal reachability graphs," in *MOBICOM'12*, 2012, pp. 377–388.

[8] V. Nicosia, J. Tang, C. Mascolo, M. Musolesi, G. Russo, and V. Latora, *Temporal Networks*, 2013, ch. Graph metrics for temporal networks.

[9] D. Braha and Y. Bar-Yam, "Time-dependent complex networks: dynamic centrality, dynamic motifs, and cycles of social interaction," in *Adaptive networks: Theory, models and applications*, T. Gross and H. Sayama, Eds. Springer, 2008, pp. 38–50.

[10] D. A. Shamma, L. Kennedy, and E. F. Churchill, "Tweet the debates: Understanding community annotation of uncollected sources," in *Proceedings of the First SIGMM Workshop on Social Media*, ser. WSM '09. New York, NY, USA: ACM, 2009, pp. 3–10.

[11] H. Kim and R. Anderson, "Temporal node centrality in complex networks," *Physical Review E*, vol. 85, pp. 1–8, 2012.

[12] P. Laflin, A. V. Mantzaris, F. Ainley, A. Otley, P. Grindrod, and D. J. Higham, "Discovering and validating influence in a dynamic online social network," *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 1311–1323, Oct. 2013. [Online]. Available: http://link.springer.com/10.1007/s13278-013-0143-7

[13] A. Alsayed and D. J. Higham, "Betweenness in time dependent networks," *Chaos, Solitons & Fractals*, vol. 72, pp. 35–48, Feb. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S096007791400229X

[14] K. Lerman, R. Ghosh, and J. H. Kang, "Centrality metric for dynamic networks," in *Proceedings of the Eighth Workshop on Mining and Learning with Graphs - MLG '10*. New York, New York, USA: ACM Press, Jul. 2010, pp. 70–77. [Online]. Available: http://dl.acm.org/citation.cfm?id=1830252.1830262

[15] E. C. Costa, A. B. Vieira, K. Wehmuth, A. Ziviani, and A. P. C. da Silva, "Time centrality in dynamic complex networks," 2015, preprint, http://arxiv.org/abs/1504.00241.

[16] V. Latora and M. Marchiori, "Efficient Behavior of Small-World Networks," *Physical Review Letters*, vol. 87, no. 19, p. 198701, Oct. 2001. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.87.198701

[17] ——, "A measure of centrality based on the network efficiency," *New Journal of Physics*, vol. 9, 2007.

[18] "Source code for the temporal centrality metrics," http://www-rp.lip6.fr/∼magnien/TemporalCentrality/.

[19] P. U. Tournoux, J. Leguay, M. Dias de Amorim, F. Benbadis, V. Conan, and J. Whitbeck, "The Accordion Phenomenon: Analysis, Characterization, and Impact on DTN Routing," in *Proceedings of the 28rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*. IEEE, 2009, pp. 1116–1124. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5062024

[20] J. Shetty and J. 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*. New York, New York, USA: ACM Press, Aug. 2005, pp. 74–81. [Online]. Available: http://dl.acm.org/citation.cfm?id=1134271.1134282

[21] S. Heymann, M. Latapy, and C. Magnien, "Outskewer: Using Skewness to Spot Outliers in Samples and Time Series," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, Aug. 2012, pp. 527–534. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6425713

[22] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *PNAS*, vol. 99, pp. 7821–7826, 2002.