Evaluating the Quality of a Network Topology through Random Walks

A.-M. Kermarrec, E. Le Merrer, B. Sericola, and G. Trédan

IRISA/INRIA - Campus de Beaulieu, 35042 Rennes, France

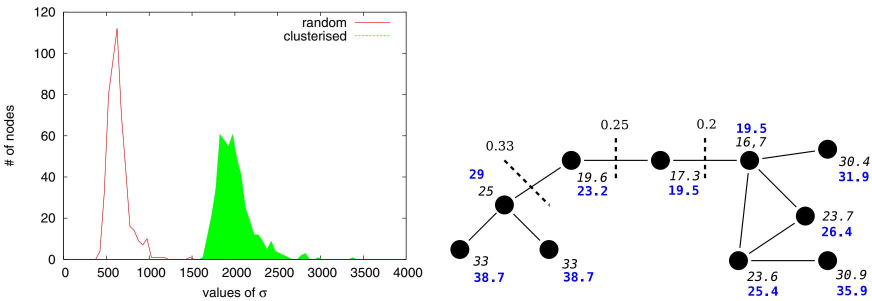
In this brief announcement we propose a distributed algorithm to assess the connectivity quality of a network, be it physical or logical. In large complex networks, some nodes may play a vital role due to their position (*e.g.* for routing or network reliability). Assessing global properties of a graph, as importance of nodes, usually involves lots of communications; doing so while keeping the overhead low is an open challenge. To that end, *centrality* notions have been introduced by researchers (see *e.g.* [Fre77]) to rank the nodes as a function of their importance in the topology. Some of these techniques are based on computing the ratio of shortest-paths that pass through any graph node. This approach has a limitation as nodes "close" from the shortest-paths do not get a better score than any other arbitrary ones. To avoid this drawback, physician Newman proposed a centralized measure of *betweenness centrality* [New03] based on random walks: counting for each node the number of visits of a random walk travelling from a node i to a target node j, and then averaging this value by all graph source/target pairs. Yet this approach relies on the full knowledge of the graph for each system node, as the random walk target node should be known by advance; this is not an option in large-scale networks. We propose a distributed solution[1] that relies on a single random walk travelling in the graph; each node only needs to be aware of its topological neighbors to forward the walk.

We consider a n node static and connected network, represented as an undirected graph $\mathcal{G} = (V, E)$, with n vertices and m edges. Each node $i \in V$, maintains its topological neighbors in $\mathcal{G}$, its degree is noted d_i.

Our approach relies on the local observation, at each node, of the variations of *return times* of a random walk on the topology. Intuitively, the more regular the visits on nodes, the more "well knitted" the network, as the walk is not periodically stuck in poorly connected parts of $\mathcal{G}$. The algorithm we propose is very simple: each graph node logs and computes the standard deviation of the return times of a permanent *unbiased random walk*, running on the topology. An unbiased random walk has a *stationary distribution* π_i, for all $i \in V$, that is $1/n$. A biased (or simple) random walk, puts mass on high degree nodes, as $\pi_i = d_i/2m$. We unbias the random walk by using the Metropolis-Hastings method [SRD$^+$06], in order to capture on nodes passage times that are only dependant of the connectivity of the graph. Every node i in $\mathcal{G}$ joins the detection process on the first passage of the walk, by creating an array Ξ_i that logs every return time. A simple solution to capture irregularity of visits on nodes is to proceed

[1] Detailed report can be found at [KMST08].

G. Taubenfeld (Ed.): DISC 2008, LNCS 5218, pp. 509–511, 2008.



(a) (b)

as follows: after the third return, a node i computes the standard deviation σ_i of the return times recorded in Ξ_i (*i.e.* the time needed for the walk, starting at i, to return to i).

Figure 1(a) presents σ values resulting on nodes after 5.10^5 walk steps, on *(i)* an *Erdös-Rényi random graph* (two nodes are connected with proba. $p = \frac{2 \ln n}{n}$) usually seen as an "healthy" graph, and *(ii)* on a *barbell graph*, consisting of two distinct cliques connected by only a link, as a model of pathological topology. $n = 1000$ in both cases. The thiner the spike, the more homogeneously the σ values are distributed on the nodes, as a result of our algorithm. The clusterized graph's values are concentrated around 4 times the average value for the random graph, indicating that visits of the walk on the nodes are far more irregular, due to the topology characteristics. Figure 1(b) plots on a micro network, with the standard deviation of random walk visits on every node (italic values), the theoretical values [KMST08] (bold), and the three lowest *conductance* values for this graph (over dashed lines). We observe that the importance of nodes in the topology is effectively correlated to the inverse of their value order: smallest σ values are effectively related to the smallest conductance values.

Distributed auto-detection of network connectivity issues could be set up, as on following examples. Once σ values on nodes allow a ranking with respect to their importance (small σ are most critical nodes), nodes may compare their σ (*e.g.* periodical *gossip* communications), to deduce vital nodes, with respect to the network topology. Another solution is for nodes to exchange passage times; suppose nodes a and b, exchange their sets Ξ_a and Ξ_b. The ratio $r_{a \to b} = \frac{\sigma(\Xi_a \cup \Xi_b)}{\sigma(\Xi_a)}$ can be exploited as a distributed cluster detector: if a and b are located in two different clusters, then the standard deviation of the union of passage dates is small, so that $r_{a \to b}$ is low. On the other side, if nodes a and b are in the same cluster, the walk is likely to hit both at very close periods, so that $r_{a \to b}$ converges to 1. a and b are thus both able to identify their respective relative position in a fully decentralized way.

Algorithm time complexity is related to graph *cover time*, as each node needs at least three visits to start an estimation process. Cover time upper bound for biased random walk is known [Fei95]: $\frac{4}{27}n^3 + o(n^3)$, for the degenerated *lollipop* graph.

References

[Fei95] Feige. A tight upper bound on the cover time for random walks on graphs. RSA 6 (1995)

[Fre77] Freeman, L.C.: A set of measures of centrality based on betweenness. Sociometry 40(1), 35–41 (1977)

[KMST08] Kermarrec, A.-M., Le Merrer, E., Sericola, B., Trédan, G.: Rr-6534 inria - evaluating topology quality through random walks (2008)

[New03] Newman, M.E.J.: A measure of betweenness centrality based on random walks (September 2003)

[SRD+06] Stutzbach, D., Rejaie, R., Duffield, N., Sen, S., Willinger, W.: On unbiased sampling for unstructured peer-to-peer networks. In: IMC 2006, pp. 27–40. ACM, New York (2006)