Dealing with Disappearance of nodes in Social Networks

Idrissa Sarr (Université Cheikh Anta Diop, Senegal)

October 2012

Outline

- Motivation
- Objectives
- Basic concepts
- Managing single node disappearance
- Managing group disappearance
- Validation
- Conclusion and future work

Analytics in 

- May help to find
  - Relevant and personalized information by following/exploring the ties between elements
    - Groups and communities within a network
    - Influential actors in a group/network
    - Network evolution and prediction
    - etc.

Analytics in 

- Knowing the groups and communities
  - Helps understand the shared interests, status, …
    - what kind of products or services to suggest?
    - How to communicate with actors?

- Identifying the influential actors
  - Can enable to gauge their impact (or influence) on the other employees or clients
  - Predict the consequences of their departure

- Understanding the dynamics of the network structure
  - allows to forecast the behavior of clients and plan new changes and products

Motivation (I)

- The network structure is very dynamic
  - entities and links appear and disappear frequently
  - Ex: social networking sites, companies, school, …
- Some nodes have more influence than others (i.e. central nodes)
  - they have the best ability to spread the information flow within the network
- The disappearance of a central node does not have the same impact as the deletion of a rather peripheral one
- Non-influential nodes, grouped as a set, may play an important role in the information flow or the network connectivity.

Objectives

- Predict the network evolution after a node/ group disappearance
  - Manage the network structure in such a way that a breakdown will never happen even if an actor set disappears
    - add new links
    - select a group/node substitute based on the leaving group configurations (or node role)

Related work

- Several studies deal with link prediction and/or network evolution in a social network
  - Liben-Nowell & Kleinberg (2003)
    - Exploit the proximity of nodes within a network to predict network structure by adding new edges.
    - Find a substitute to the deleted node provided it plays a role in the network. New links can then be established according to some predefined options (e.g. clique, ring).
  - Missaoui et *al.* (2011) Sarr et Missaoui (2012)
    - predict the new structure of a social network once a node disappears
    - possibly find a substitute for the leaving node provided it plays a role in the network and/or the information flow quality.
- Most of the proposed solutions
  - can explode the number of links within the network.
  - do not deal with a group disappearance that can happen in any social network

Basic concepts

Some definitions

- Degree centrality
  - provides the number of direct links of a node
    - it helps identify leaders which have the highest number of links within the network.

- Betweenness centrality
  - expresses the amount of control that a node (or a group of vertices) possesses over the interactions of other nodes in the network.
    - It is high for mediators (or brokers) which are nodes that act as intermediaries between other nodes or as joins between communities.

- Closeness centrality
  - indicates how a node is close to the other nodes in the network and hence how fast information circulates from a given node to other reachable nodes in the network.

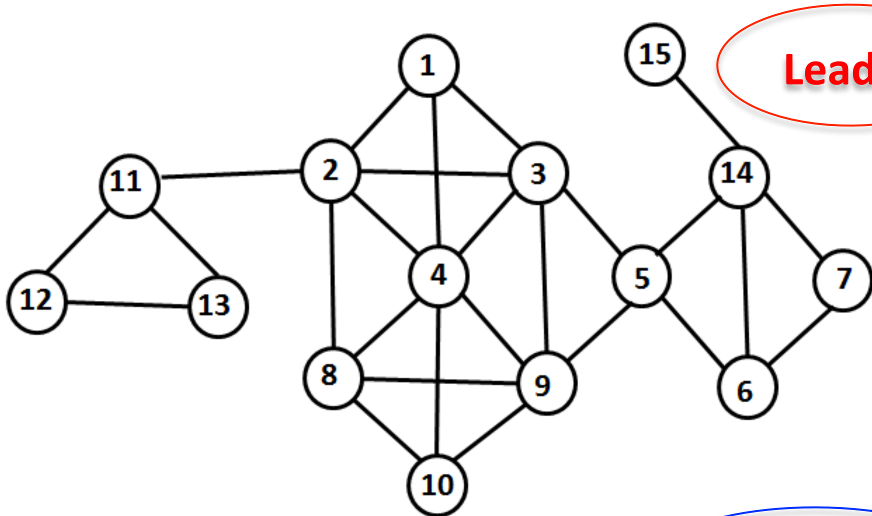
Node Classes

- Nodes inside a group are classified into two categories based on their centrality measures (degree, betweenness, and closeness centrality)
  - Critical nodes
    - Nodes playing central roles: i.e. nodes with the highest centrality measures
    - Three types of critical nodes : *leader*, *mediator* and *witness*
    - critical disappearance : critical nodes that are unique in their class
  - Non critical nodes
    - Nodes playing less central roles
    - Two possible types
      - ***Finger*** : node whose centrality measure deviates slightly from the one of a critical node
      - ***Follower*** : neither critical nor finger node
    - non-critical disappearance : either non-critical deleted node or a critical node which is not unique in its class

Node Classes

Witness

Leader



	Degree centrality	Betweenness centrality	Closeness centrality	Eccen-tricity
1	0.214	0	0.424	4
2	0.357	0.371	0.482	4
3	0.357	0.322	**0.518**	**3**
4	**0.428**	0.067	0.466	4
5	0.285	**0.439**	0.482	4
6	0.214	0.06	0.368	5
7	0.142	0	0.285	6
8	0.285	0.036	0.437	4
9	0.357	0.146	0.466	4
10	0.214	0	0.388	4
11	0.214	0.263	0.368	5
12	0.142	0	0.280	6
13	0.142	0	0.280	6
14	0.285	0.203	0.378	5
15	0.07	0	0.280	6

Mediator

If we set a deviation of 20%, then node 2, 3 and 9 will be the *fingers* of the leader node 4.

Information Flow Quality

- Hypothesis
  - Assume *S* a connected, undirected and unweighted network

- Eccentricity of a node *i* : $\varepsilon\ (i)$
  - Greatest geodesic distance between *i* and any other node *j* in the network

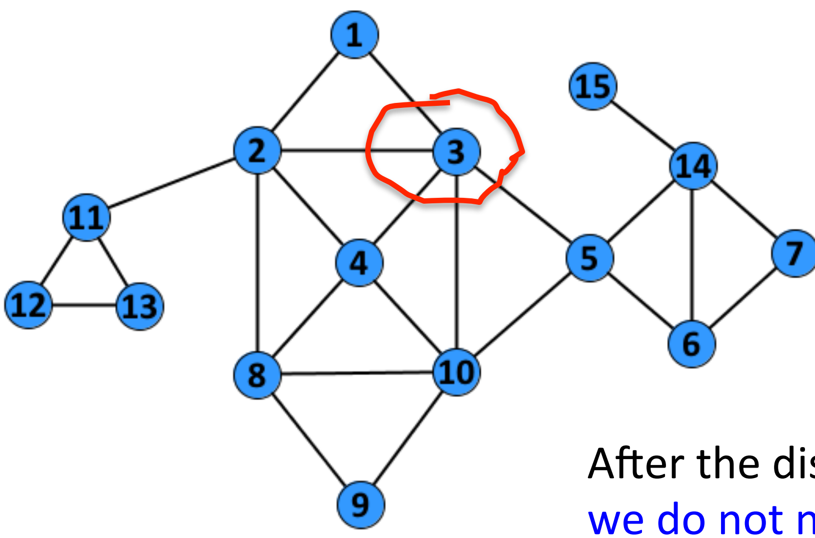
- Information flow degree of S : $\lambda_F\ (S)$
  - Total number *N* of nodes that can receive the information

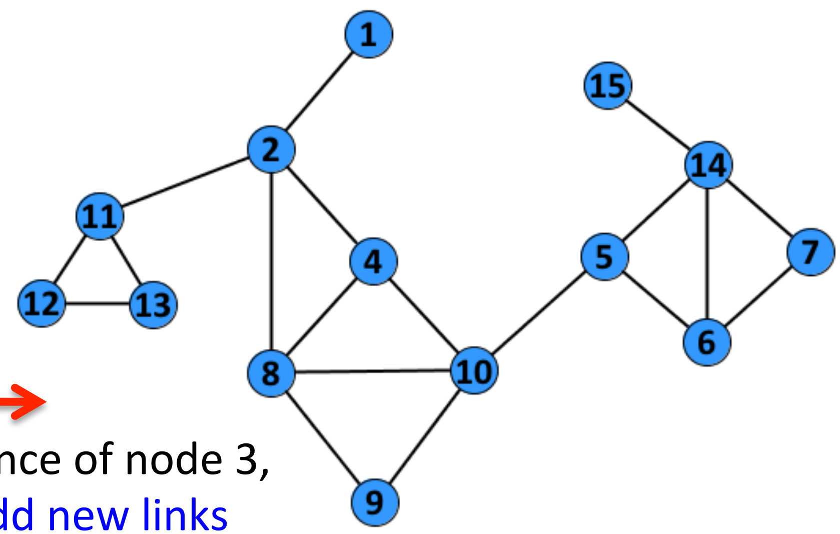
– Information flow (IF) quality of S : $Y_F\ (S)$
  - $Y_F\ (S) = \ < \lambda_F\ (S),\ \varepsilon\ (w) >$
  - $\varepsilon\ (w)$ is the eccentricity of the witness that is the best entry of the network based on its closeness centrality

Dealing with single node disappearance

Managing a single node disappearance?

- Main idea
  - After the disappearance of a node, add new links and possibly find a substitute if the information flow quality is not ensured (e.g., disconnected network, delays in information propagation)

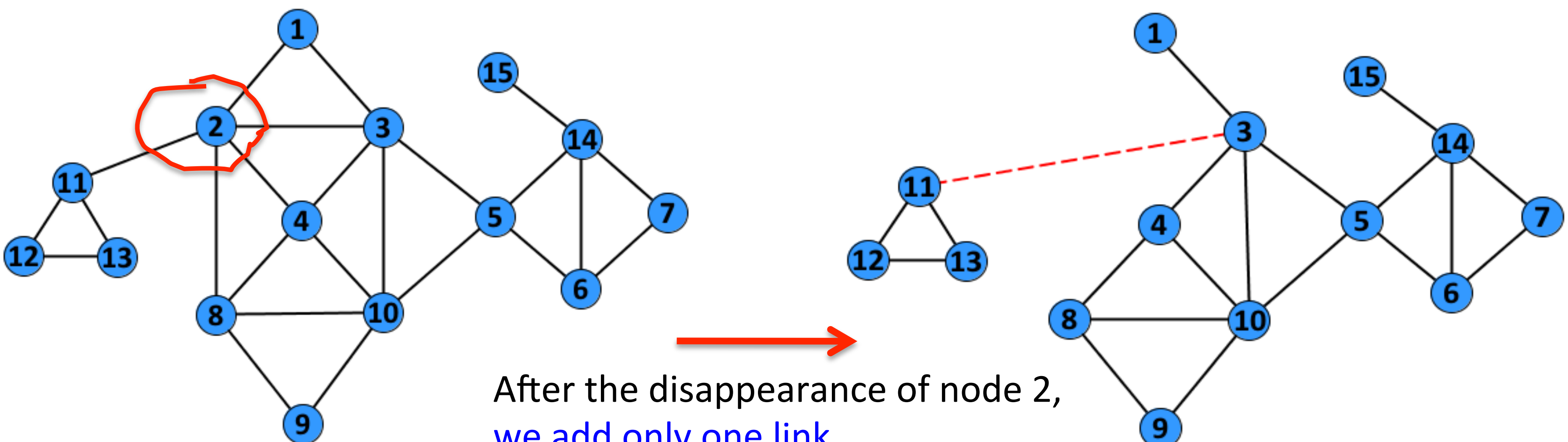




After the disappearance of node 3,
we do not need to add new links
since the network is still connected

Managing a single node disappearance?

- Main idea
  - After the disappearance of a node, add new links and possibly find a substitute if the information flow quality is not ensured
    - i.e. if the network is disconnected



After the disappearance of node 2,
we add only one link

Requirements

- Let S' be the network obtained after
  - (i) deleting node Ni from S,
  - (ii) removing links attached to Ni,
  - and (iii) adding new links.
- The idea is to minimize the number of added links while maintaining the following compound condition:

$$\frac{\lambda_F(S) - \lambda_F(S')}{\lambda_F(S)} \leq \varepsilon_1 \quad \wedge \quad \frac{\epsilon_{S'}(w) - \epsilon_S(w)}{\epsilon_S(w)} \leq \varepsilon_2$$

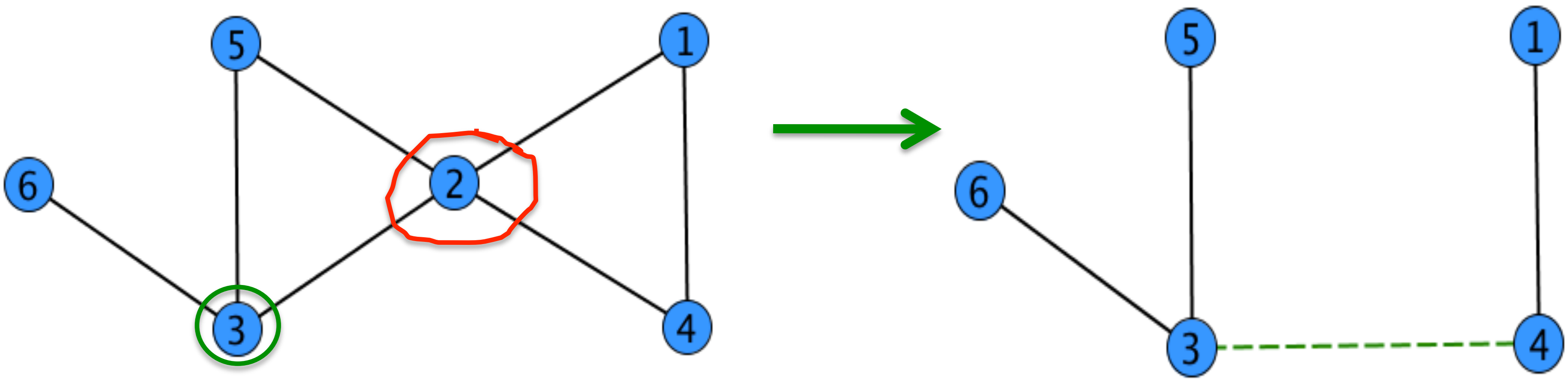
In other words, after a node disappearance, what is the "optimal" update that maintains the IF quality *w.r.t* some given deviations ε_1 *and* ε_2

Dealing with critical disappearance

- A critical disappearance of *Nc* is handled as follows:
  1. Nc has some fingers (at least one). In this case, the first finger (F1) (finger with the highest centrality measure) is chosen as the substitute
  2. There is no finger to replace the leaving critical node. Then, one of the remaining critical nodes is selected as a substitute
     - The critical node with the shortest geodesic distance to Nc is used as a substitute
  3. There is no other critical nodes, which means that the leaving node was the only critical node (e.g., a central node of a star network). Then, one of the Nc neighbors with the highest centrality measure is selected as the substitute.

- For each one of the three cases, any neighbor of Nc is linked to the substitute if ever it has not been already linked

Critical disappearance – JOAN-C



Node 3 is the substitute since it may be the finger and/or the neighbor with the highest centrality measure

Dealing with non-critical disappearance

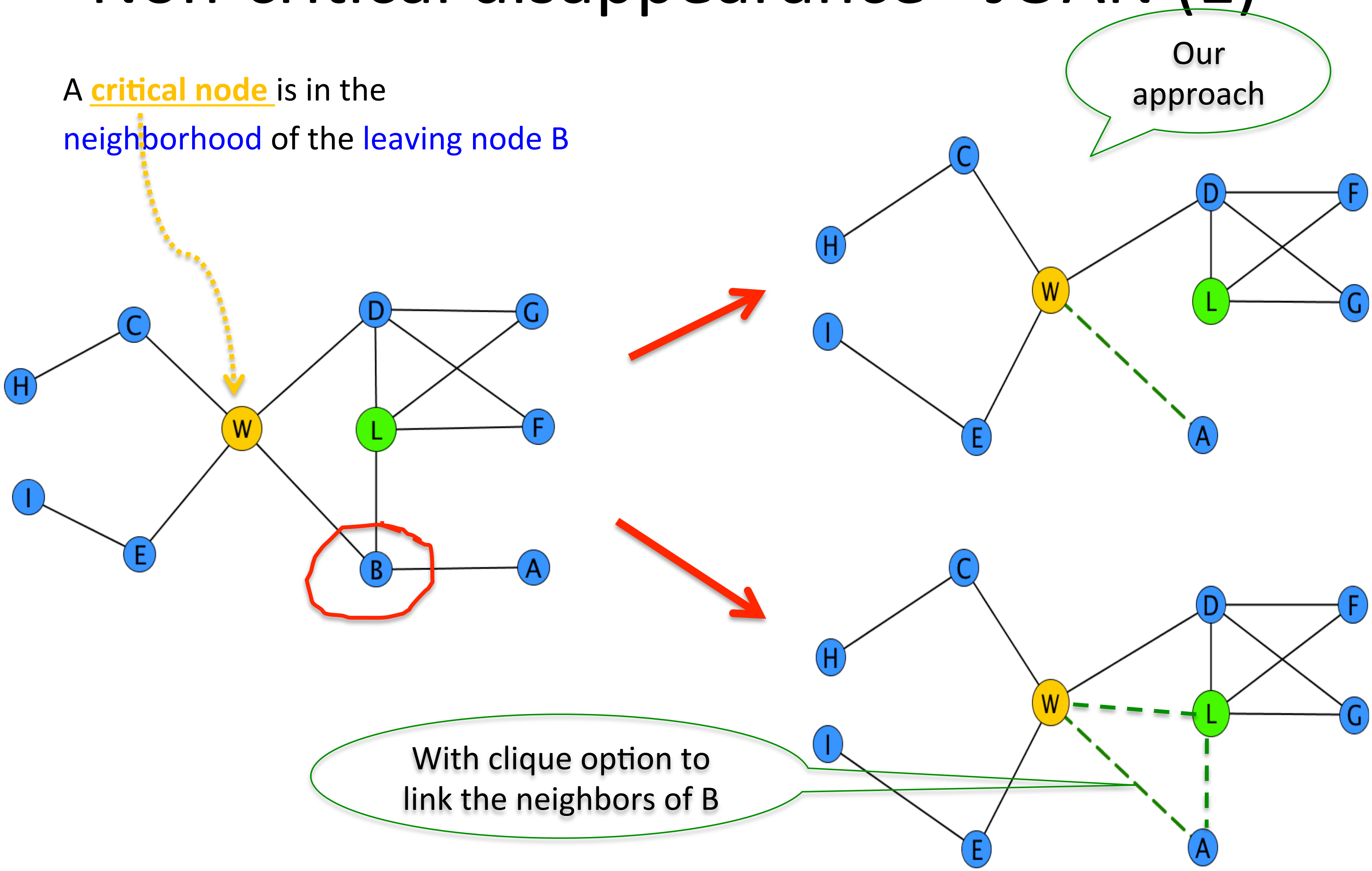
- A non-critical disappearance of *Ni* is handled as follows:

1. The degree of *Ni* is equal to one, i.e., *Ni* has one neighbor.
   - Nothing to do because both conditions in the previous Formula are true for a network of size at least equal to $1/\varepsilon_1$.

2. The degree of Ni is greater than one.
   - While the tolerated information flow degree is not ensured, a link between two neighbors *Nj* and *Nk* of *Ni* is added as follows:
     1. A direct link is added if *Nj* or *Nk* is a critical node.
     2. An indirect link (between *Nj* and *Nk*) is added if neither *Nj* nor *Nk* is a critical one. The indirect link is set between *Nj* and *Nk* through another neighbor *Ns* which has the shortest path to a critical node *Nc*.

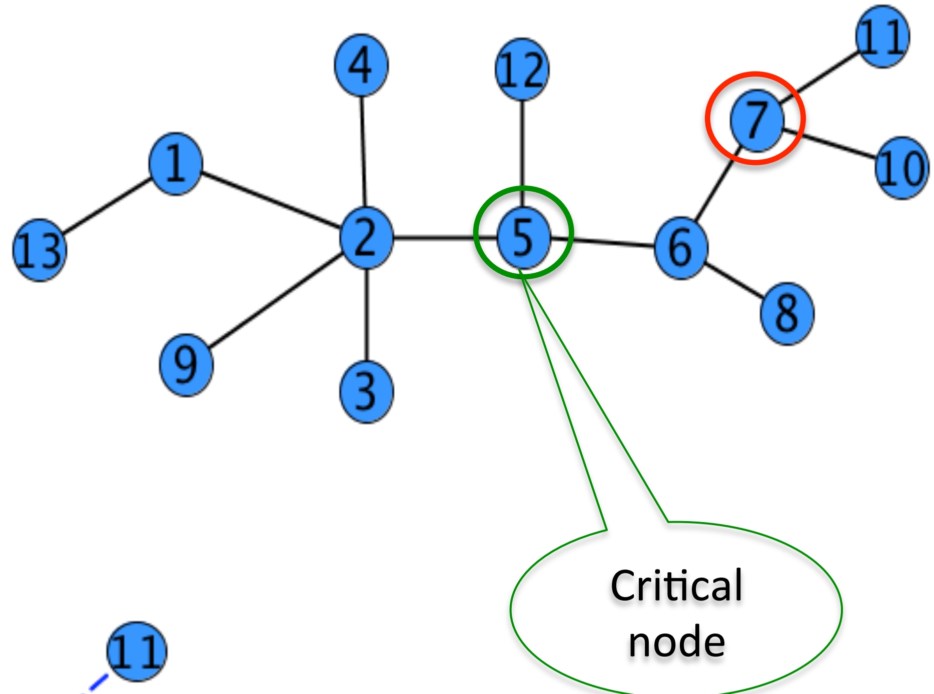
Non-critical disappearance - JOAN (1)

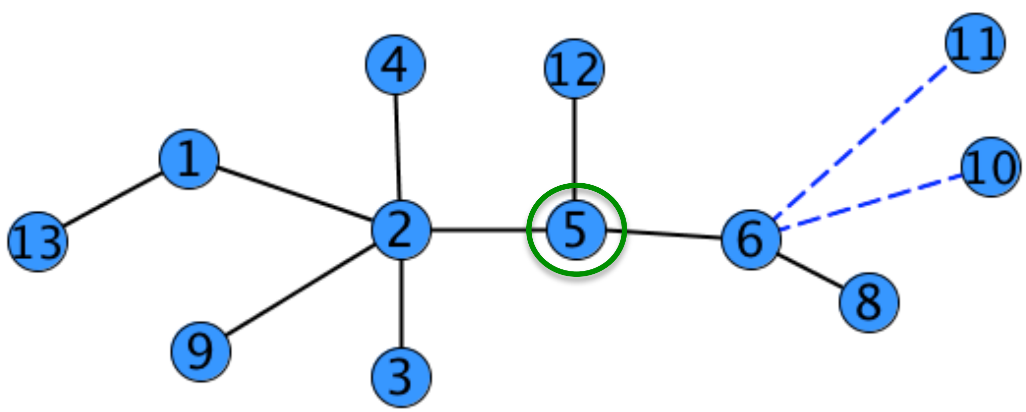
A critical node is in the neighborhood of the leaving node B



Non-critical disappearance - JOAN (2)

A critical node is not in the neighborhood of the **leaving node 7**



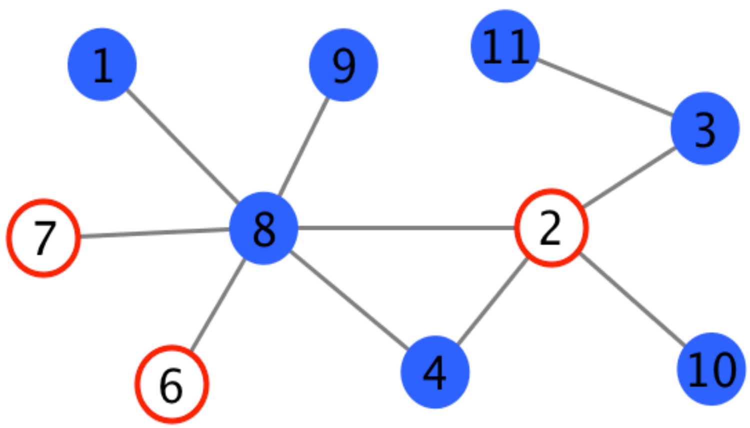


Managing a group disappearance

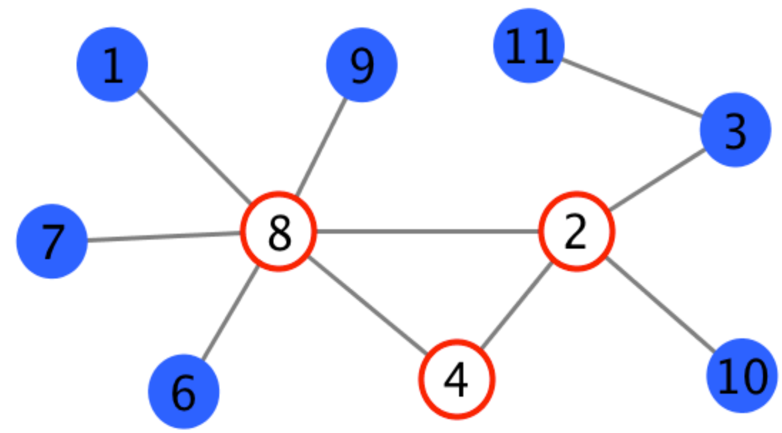
Configurations of a group

- A group G, can have one of the 3 configurations
  - Scattered
    - there is no direct link for any couple of nodes in G
  - Continuous group
    - there exists a direct link for any couple of nodes (Ni, Nk)
    - There exists an indirect link between Ni and Nk through only nodes inside G.
  - Hybrid group
    - Combination of both previous configurations coexist

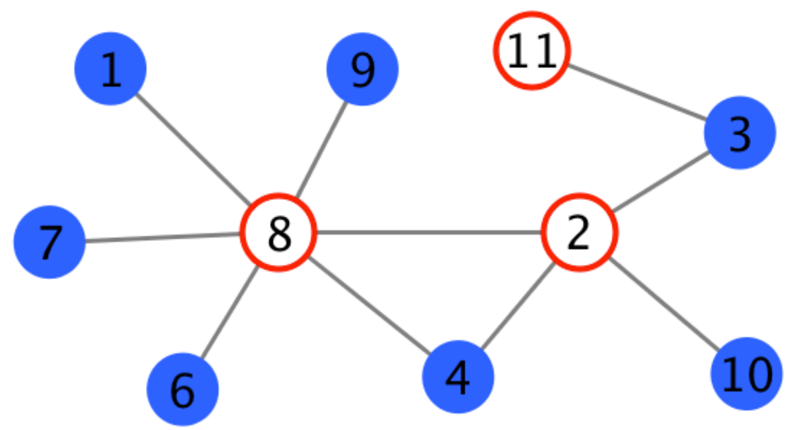
Example



scattered



grouped



hybrid

Scattered group

- each node N_i in G is far from the other ones
  - its disappearance will not involve another node N_k in the group.
    - i.e. by using *JOAN* or *JOAN-C*, we can face the disappearance of N_i without using any $N_k \in G$ as a substitute or linking directly N_i to N_k.
- For each node $N_i \in G$,
  - apply the *JOAN* or *JOAN-C* algorithm based on the class of N_i to manage its disappearance.

Contiguous group

- disappearance of a contiguous group is handled in one shot.
  - JOAN or JOAN-C not efficient
    - may lead to tie two nodes that will leave afterwards because they belong to the same leaving group

- Every leaving group is replaced by another one to:
  - preserve the overall shape of the network
  - and ensure the information dissemination

- To replace a group G, we consider all neighbors Γ(G) of any node in G.
  - Let Γ(Ni) be the set of neighbors of a node Ni

$$\Gamma(G) = \{\bigcup_i \Gamma(N_i) \setminus G \mid N_i \in G\}$$

Contiguous group : common neighbors

- To know the nodes to consider first when finding the substitute of a leaving group, we define common neighbors.

- Total common neighbors *$\Gamma_T(G)$*
  - the set of nodes that belong to each neighborhood of a node in *G*

$$\Gamma_T(G) = \{\bigcap_i \Gamma(N_i) \mid N_i \in \Gamma(G)\}$$

- Partial common neighbors *$\Gamma_P(G)$*
  - The union of the intersections of the neighbors of any couple of nodes in G.

$$\Gamma_P(G) = \bigcup_{i,k} \{\Gamma(N_i) \cap \Gamma(N_k) \mid (N_i, N_k) \in \Gamma(G)\}$$

Finding a substitute

- Let *G'* = substitute of a leaving group *G*

- $\Gamma_T(G) \neq \varnothing$: there is at least one common neighbor for all nodes in *G*
  - Γ_T (G) is used as a substitute group

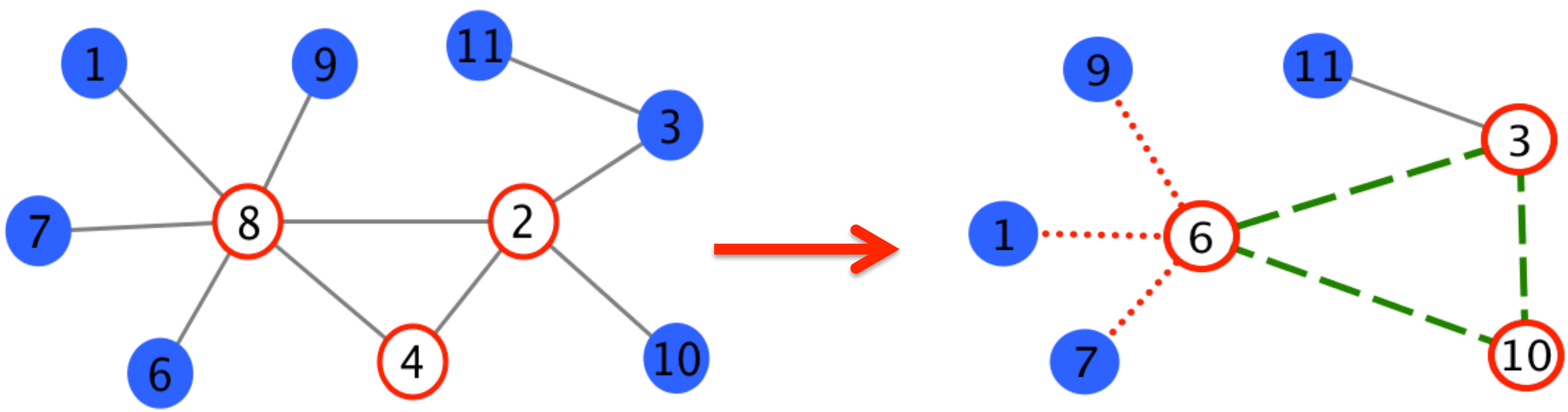
- $\Gamma_T(G) = \varnothing \wedge \Gamma_P(G) \neq \varnothing$
  - Set G' by choosing *n* nodes in $\Gamma_P(G)$ in such a way that
    - $|G| = |G'|$
    - G' has the highest group degree centrality

- $\Gamma_T(G) = \varnothing \wedge \Gamma_P(G) = \varnothing$
  - Set G' by gathering a set of nodes from $\Gamma(G)$ in order to get a centrality value similar to or higher than the group degree centrality of G

Connecting G’ to G neighborhood

- After finding G’:
  - add direct links between any couple of nodes in G’ if ever they are not yet directly or indirectly linked.
  - for any node $N_i \in \Gamma(G)$ and $N_i \notin G'$, we add a link between N_i and only one of the nodes in the substitute group G’.

Example



Other group configurations

- Hybrid
  - first identify the scattered and contiguous subgroups,
    - for each subgroup, we apply one of the procedures presented earlier.

- Group disappearance with identified communities
  - a leaving group can be localized either in one single community or spread over many communities
    - When the group is distributed over several communities, we first identify such communities and the corresponding subgroups.
      - For each subgroup within a community, we apply the procedure based on its nature (e.g., scattered)

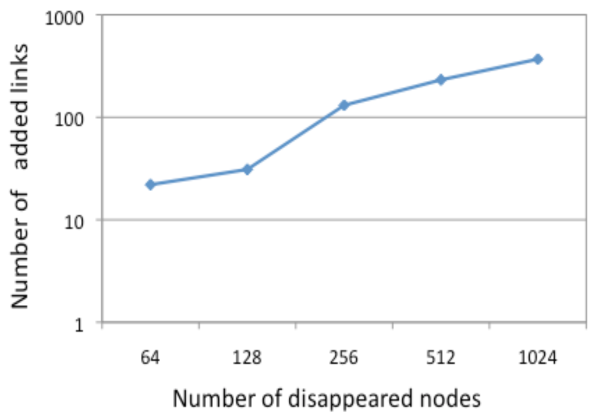
    - When the overall group of the leaving node belongs to the same community, apply the procedure described above for a subgroup

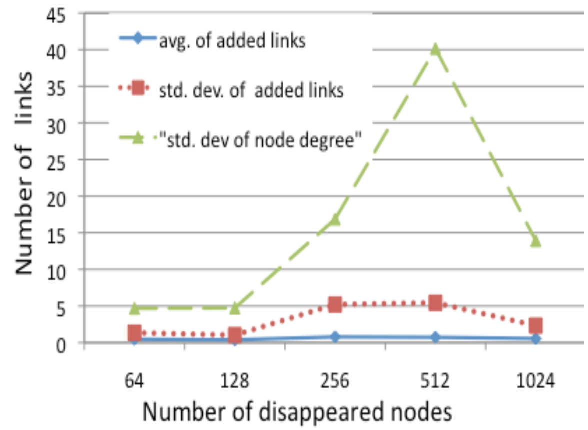
Validation

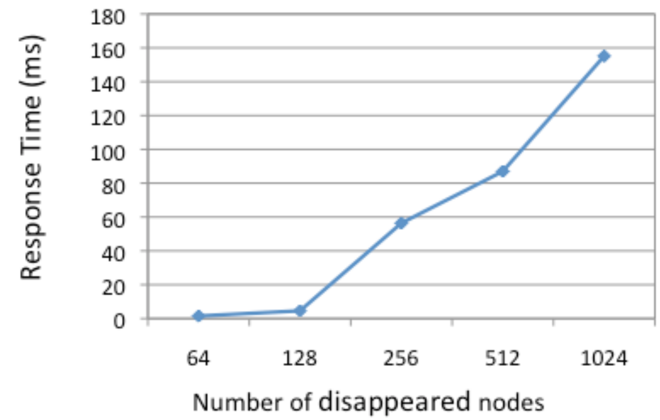
- Goal
  - Evaluate the performance in terms of execution time and number of added links when node disappearance becomes more frequent and group configurations vary.
- Data sets
  - Autonomous systems AS-733 of the Stanford University dataset (2009)
    - Undirected and unweighted network with 6474 nodes and 13233 edges.
    - describes a graph of Internet routers with a communication network model of who-talks-to-whom
- Platform
  - A prototype that uses NetworkX and the Python language.
  - Intel Core i5 with 8 GB of RAM and 3.20 GHz running under Linux Ubuntu
  - cloud infrastructure PiCloud

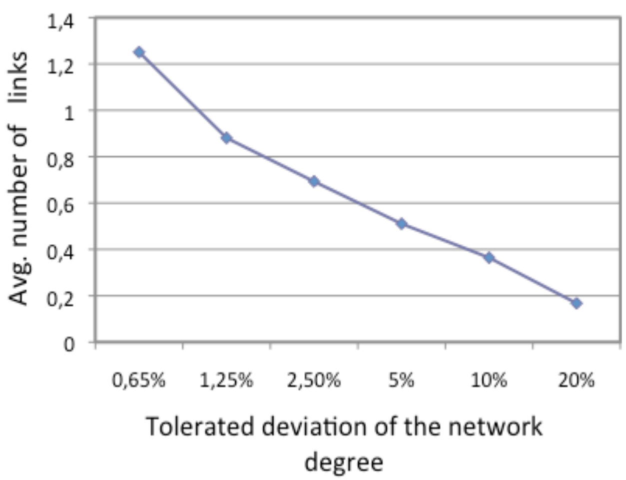
Single node disappearance results

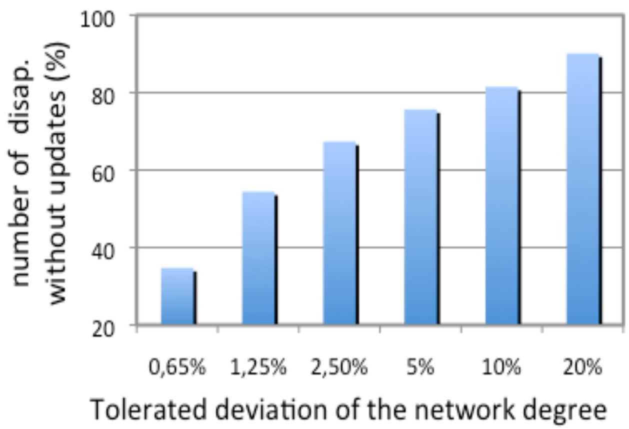
- Overall performance
  - For a deviation = 0 and sparse graphs
    - Average added links : 5 links per disappearance
    - Average response time : less than 80 ms
- Impact of the information flow deviation
  - beyond a deviation of 2,5%
    - Average added links is less than 1 per disappearance
    - The eccentricity decreases at least by a factor of 60%
    - For more than 60% of deletions, no network update because the requirements are already ensured.

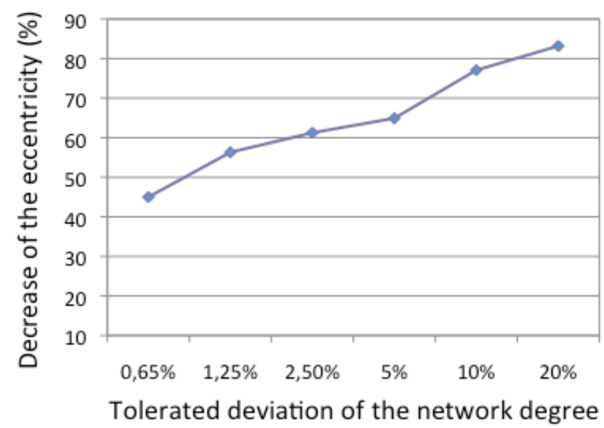










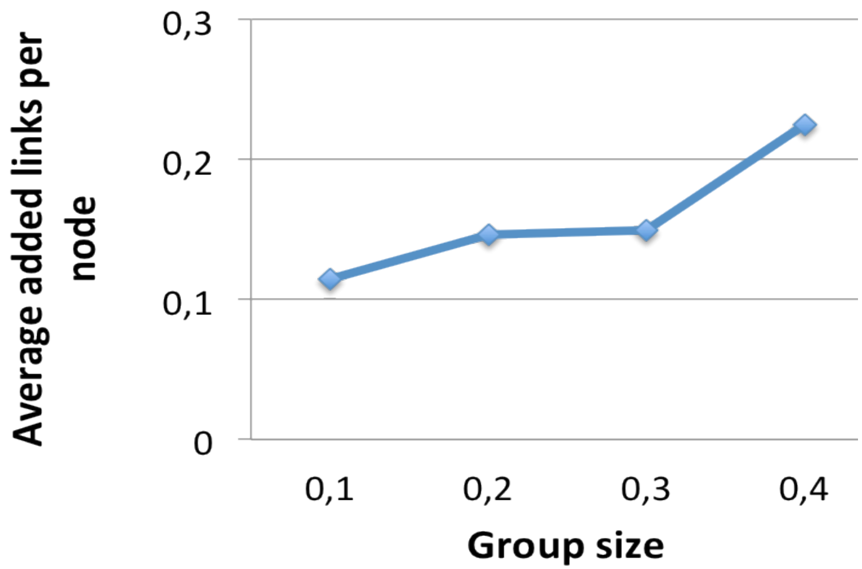


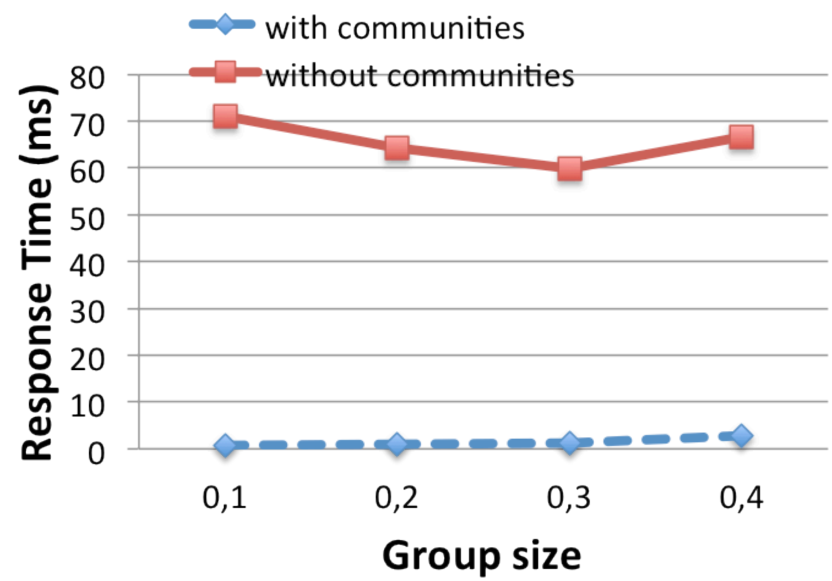
http://larim.uqo.ca/SNA/SarrSNAM.pdf

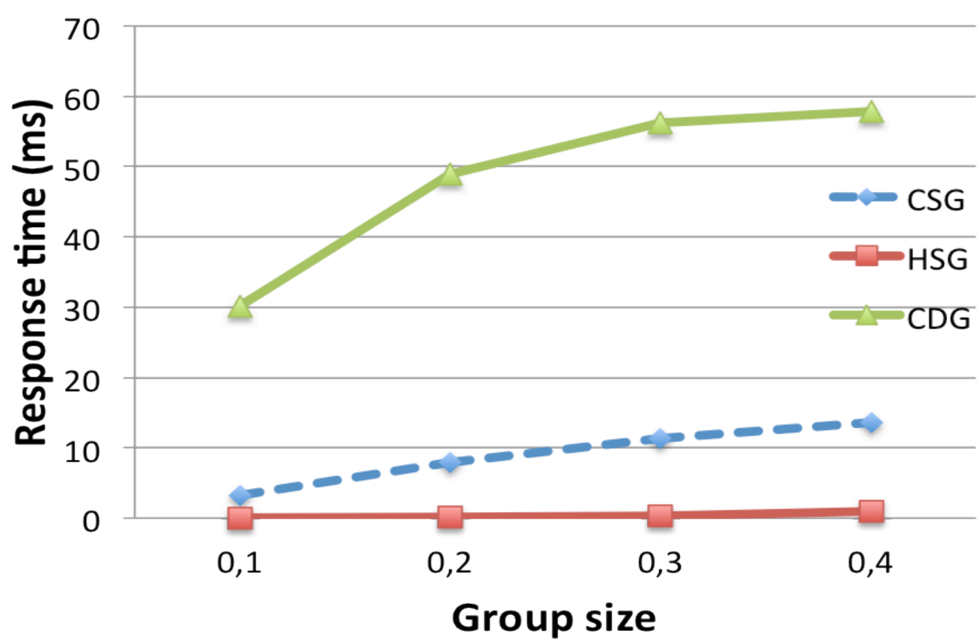
Group disappearance results

- Overall performance
  - Low increase of the number of added links
    - Average added links : 0,2 links per disappearance with 40% of disappearance.
  - Response time depends on the network density:
    - sparse graph (with hybrid group) leads to less added links and is less time consuming than dense graph.
  - Community identification gives more performances
    - Communities lead to parallelism and require less than 5ms where lack of communities consume 50ms to update network after a group disappearance .

Validation graphs







Future work

- Ongoing work:
  - we plan a more pro-active to notify and recommend ways to strengthen the weak links.

- Consider additional features such as the intensity of interactions among nodes and the direction of links to better capture the role of nodes in information dissemination

- Characterize influential actors in the network based on the roles we defined earlier (e.g., a witness and its top ranked fingers).