Additional Materials A: GCC and Robustness

This appendix contains further investigations on the GCC (Giant Connected Component) evolution and also some preliminary results on a robustness attack we performed in our wire transfer network. These experiments are mentioned in Sect. “Hubs and connectivity” of the paper.

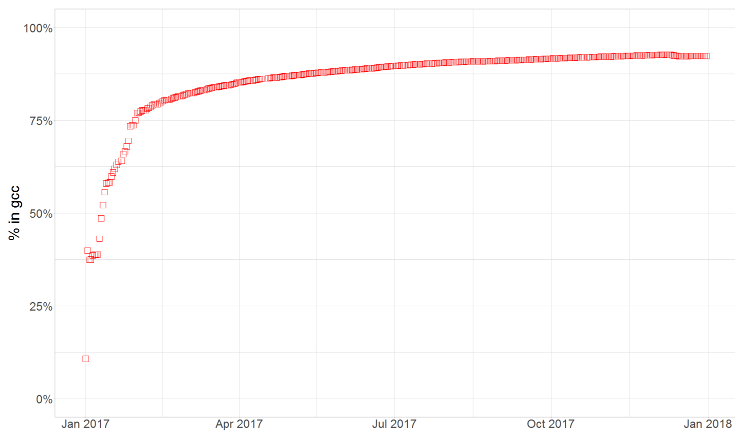

Figure A1: Giant component evolution over time: each dot in the figure represents the daily percentage of the nodes connected in the largest connected component.The percentage of nodes into the GCC grows steadily, with no spikes, in a slow process across two years.

Fig. A1 shows how the percentage of nodes in the GCC grows day by day during 2017. Among the highest degree nodes (first percentile), 94.5% are in the GCC since the first month of observation, that is January 2017. Their role in the connectivity of the network is clearly highlighted also by the network’s robustness, i.e., its ability to resist against random failures (usually defined as *node removal*). Intuitively, removing a hub can create a larger damage to the connectivity of the network respect to failures that can occur at a random node.

Fig. A2 shows how easily the network disconnects if a removal attack is targeted against different types of nodes. For instance, when the targets are the highest degree nodes (red), removing only the top 1% nodes from the giant component, will make the GCC size decrease more than 50%, and with a removal of 15% of the nodes the network falls apart in a set of very small and isolated

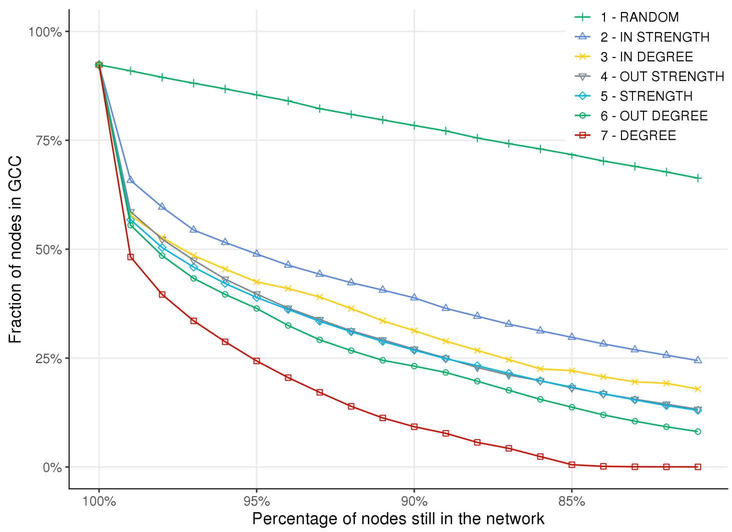

Figure A2: Robustness attack: the network disconnects if a removal attack is targeted against hubs by degree (red line - squares), rather than against random nodes (dark green line - crosses): removing only the top 1% nodes into the giant component, the GCC size is decreased more than 50%. Removing nodes by in and out degree, strength, in and out strength shows the same behaviour exhibited by degree removal, though in a softer way.

components. On the contrary, a random attack is less effective and even if 20% of the nodes are removed, the largest connected component resists including more than 70% of the remaining nodes. Removing nodes by in and out degree, strength, in and out strength shows the same behaviour exhibited by degree removal, though in a softer way.

Although the quick disruption of the connectivity of the network after a hub removal attack is a pretty predictable outcome, it is not clear if every network with the same size and degree distribution would behave the same way, or if this behaviour is particularly enhanced or diminished by the exact configuration of ISP network. To address this question, we compared the robustness of ISP network against a hub removal attack with:

- a random node removal attack
- a hub removal attack on a degree-preserving randomization of the network
- a random node removal attack on a degree-preserving randomization of the network

Results are displayed in Fig. A3. When removing random nodes in both networks, the connectivity slowly falls in a similar way; however, Intesa Sanpaolo's network's connectivity seems to be more sensitive to a hub removal attack, as the GCC tears apart faster than in the randomized network. It seems a signal that hubs in Intesa Sanpaolo's network are crucial to link parts of the network that otherwise would be disconnected, beyond what happens in same size and degree distribution networks.

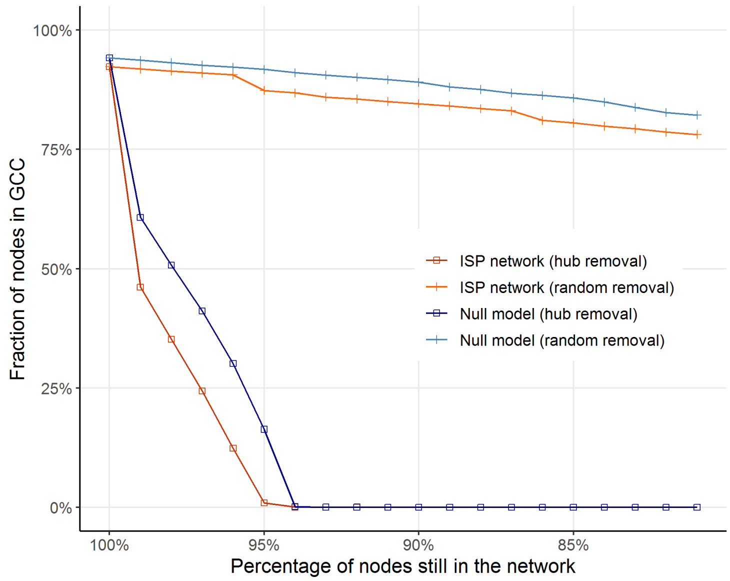

Figure A3: Robustness attack (ISP vs random): the Intesa Sanpaolo's network (red squares) disconnects faster than a degree-preserving randomization of the network (blue squares) if a removal attack is targeted against hubs by degree. Targeting random nodes leads to similar results in both networks: connectivity slowly disrupts.

In these experiments it might be interesting to also include the standard betweenness based attacks. This analysis would require greater computational power than the one we had during our research: for internal policies, it was not possible, for privacy reasons, to upload the data to a high performance computing facility installed in remote machines, and such HPC system was not available within Intesa Sanpaolo Innovation Center. Also, the interpretation of such connectivity result is not trivial (which is the effect of a wire transfer network relying on few highly connected hubs?), and further investigation would be needed to understand and exploit the outcomes of these analyses.

As a confirmation that legal entities are in charge of keeping the network connected, however, we enquired how the connectivity of the network would change if it was a natural persons only network. We extracted a projection of the whole network by simply removing legal entities nodes and edges, and computed number and size of the several components the network splits into. Fig. A4 shows a comparison between the distribution of the sizes of the components in the whole network and in the natural persons only network.

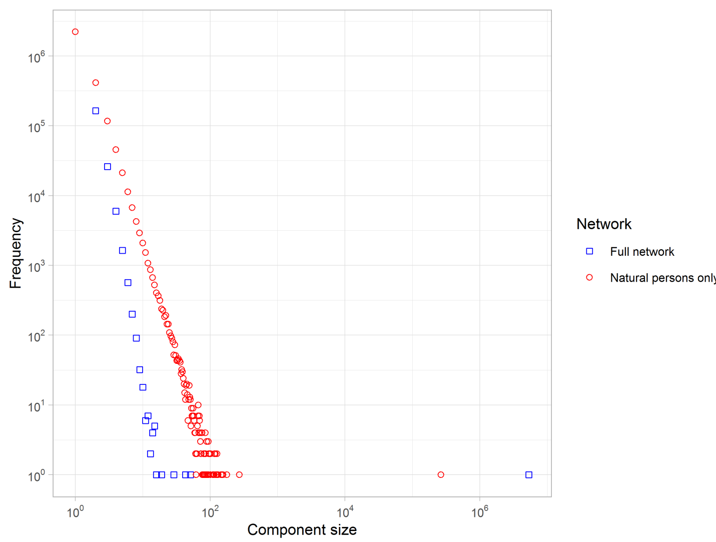

Figure A4: Sizes of components in the natural persons network: red circles show the number of x-size components in the natural persons network, blue squares shows the number of x-size components in the whole network. We can appreciate three effects when removing legal entities: GCC gets smaller, $\sim 200k$ nodes against 5.5 millions; 2.2 million nodes disconnect (component size = 1); many components with few to hundreds nodes stay disconnected from the GCC.

If we remove legal entities from the network, it falls apart into several smaller disconnected components. We can appreciate three main effects:

- The GCC shrinks to $\sim 200k$ nodes, against 5.5 million nodes in the original network. As we pointed out in the paper, 73% of the original nodes are natural persons, hence this projection of the network still keeps 73% of nodes. The GCC shrinkage is not fully explained by the nodes number reduction alone.
- 2.2 million nodes disconnect: there is a red circle pointing to component size = 1. There were no disconnected nodes in the original network. 2.2 million is more than half of the natural persons only networks' nodes.
- There are way more middle-size disconnected components, some with more than 100 nodes.