1 Communities Detection

In order to test our hypothesis we ran a community detection algorithm on the two networks under analysis: the one produced by the toy model and the Twitter one. We used the algorithm developed by Blondel at al. [1], which is found to be one of the best algorithms for community detection based on modularity optimization, as shown by Fortunato in his review of communities detection methods[2]. The algorithm works in two phases: in the first one the modularity is optimized by allowing only local changes of communities, and in the second one the found communities are aggregated in order to build a new network of communities. These two phases are repeated iteratively until no increase of modularity is possible. In the future we plan to expand the analysis by using also other algorithms, such as Infomap[3], and compare the obtained results. The analysis was carried out in two steps: first we applied the algorithm to the binary unweighted network, considering in this way only the *structural* relationships between the nodes. These are given, in the case of the real data, by looking at who is *following* who on Twitter, which is an explicitly declared relationship. The limitation of this method is that most Twitter users follow a very large number of people, but in practice they see and share only the tweets belonging to a relatively small subset of the users they follow. Therefore the second step was to apply the algorithm to a weighted version of the networks, that we obtained by weighting each structural link with the value of the mutual information between the two users. The results of this analysis are shown in Table 1.

	Synthetic Network		Twitter Network	
	Binary	Weighted	Binary	Weighted
Number of Detected Communities	8	4	26	37
Optimal Modularity Value	0.264	0.625	0.260	0.404

Table 1: Communities Detection Results

We can observe a clear increase in the optimal value of modularity when we consider the weighted network, which indicates that the network has a more defined community structure. In the case of synthetic data the value triples, whereas in the case of real data it doubles. This result shows, as we expected, that if we weight the structural links of a network with a measure of the communication flow among the nodes, giving thus more importance to the links between two users that actually share information and less to the links that are there but are not exploited, we can detect a finer community structure. In particular, for what concerns the synthetic data, we can observe that in the weighted case we find exactly four communities, which is coherent to the way the toy model was built. A visual representation of the results is given in Figure 1 for the synthetic data and in Figure 2 for the Twitter data.
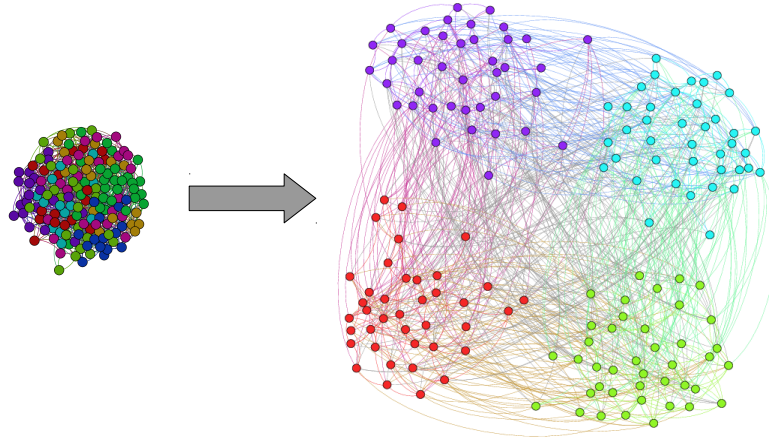
Figure 1: Visualisation of the synthetic network obtained using binary edges (left figure) and weighted ones (right figure). Node colors indicate the community affiliation.
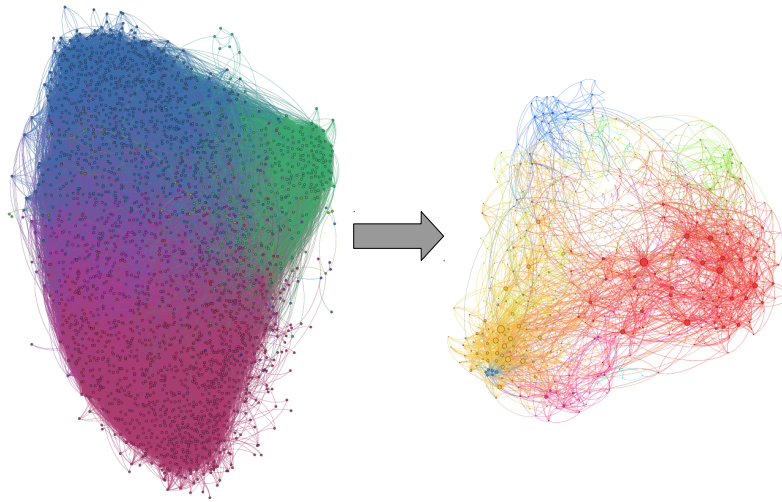

Figure 2: Visualisation of the Twitter network obtained using binary edges (left figure) and weighted ones (right figure). Node colors indicate the community affiliation.

2 Sanity Check

Dave, could you please take care of this part since it's you who worked on that?

References

[1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[2] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[3] Martin Rosvall and Carl T Bergstrom. Mapping change in large networks. *PloS one*, 5(1):e8694, 2010.