

Towards Democratic Group Detection in Complex Networks

Michele Coscia, Fosca Giannotti, Dino Pedreschi

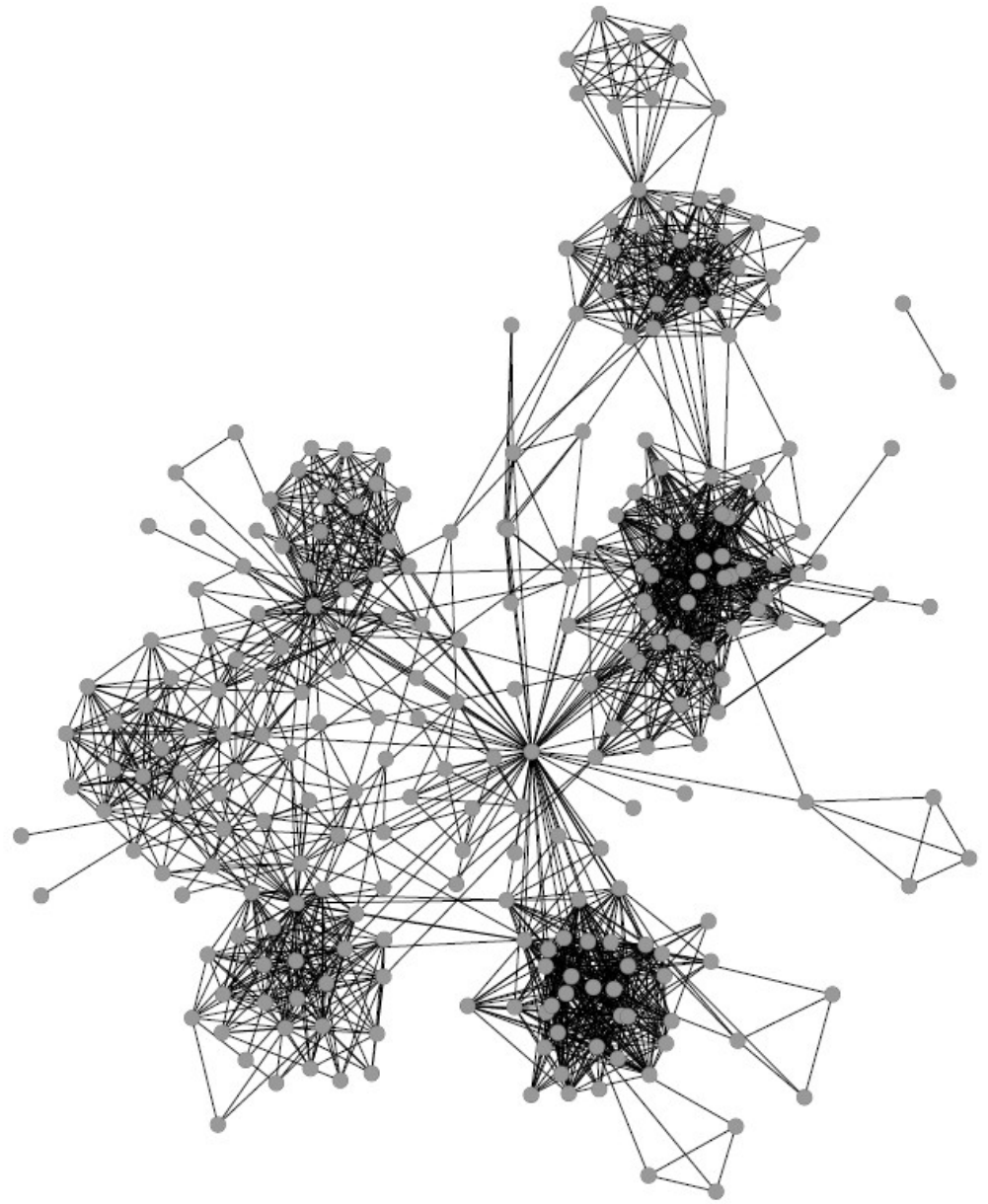
Computer Science Dep., University of Pisa, Italy {coscia,pedre}@di.unipi.it
ISTI - CNR, Area della Ricerca di Pisa, Italy fosca.giannotti@isti.cnr.it

International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction

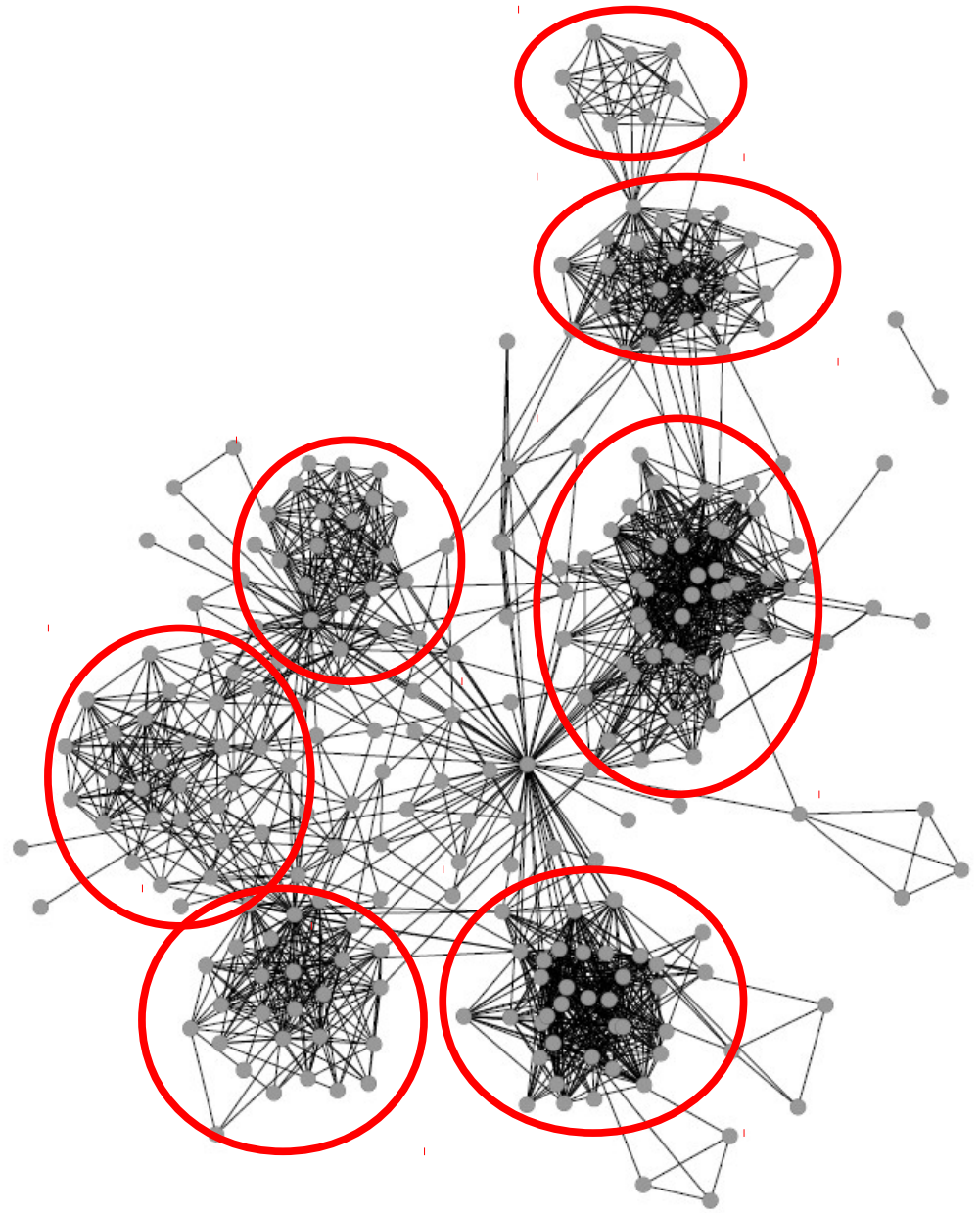
April 4th, 2012



Communities

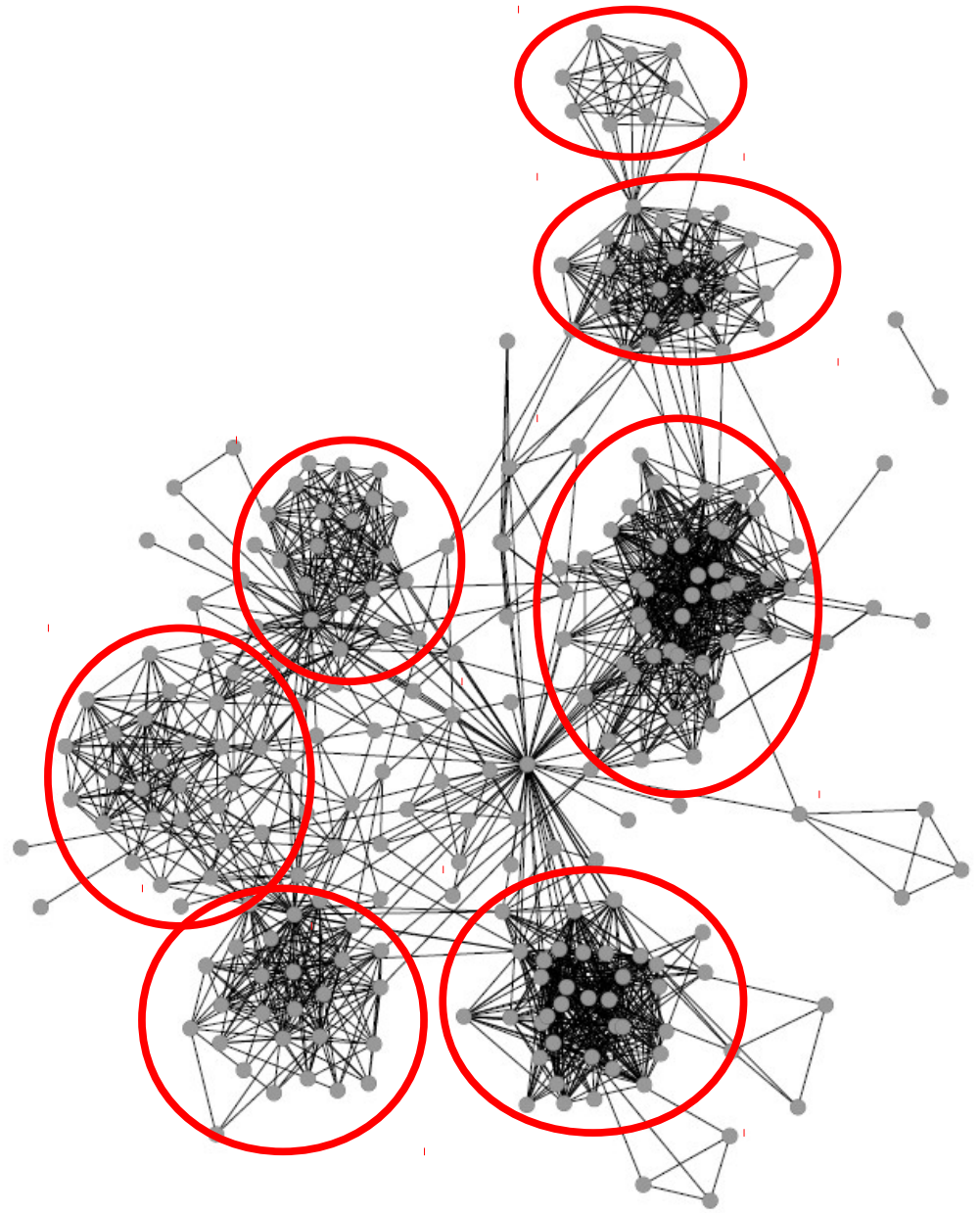


Communities

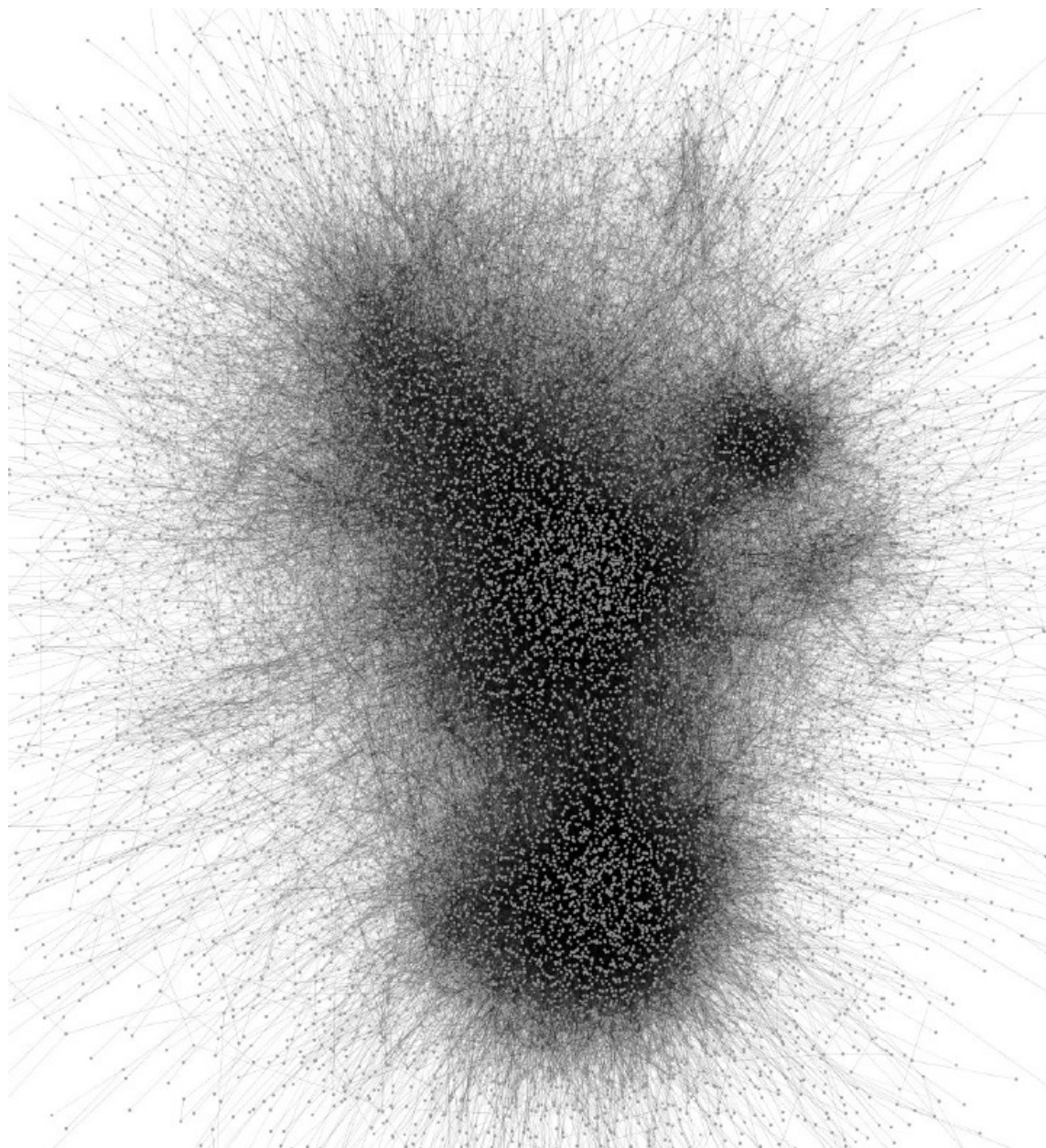


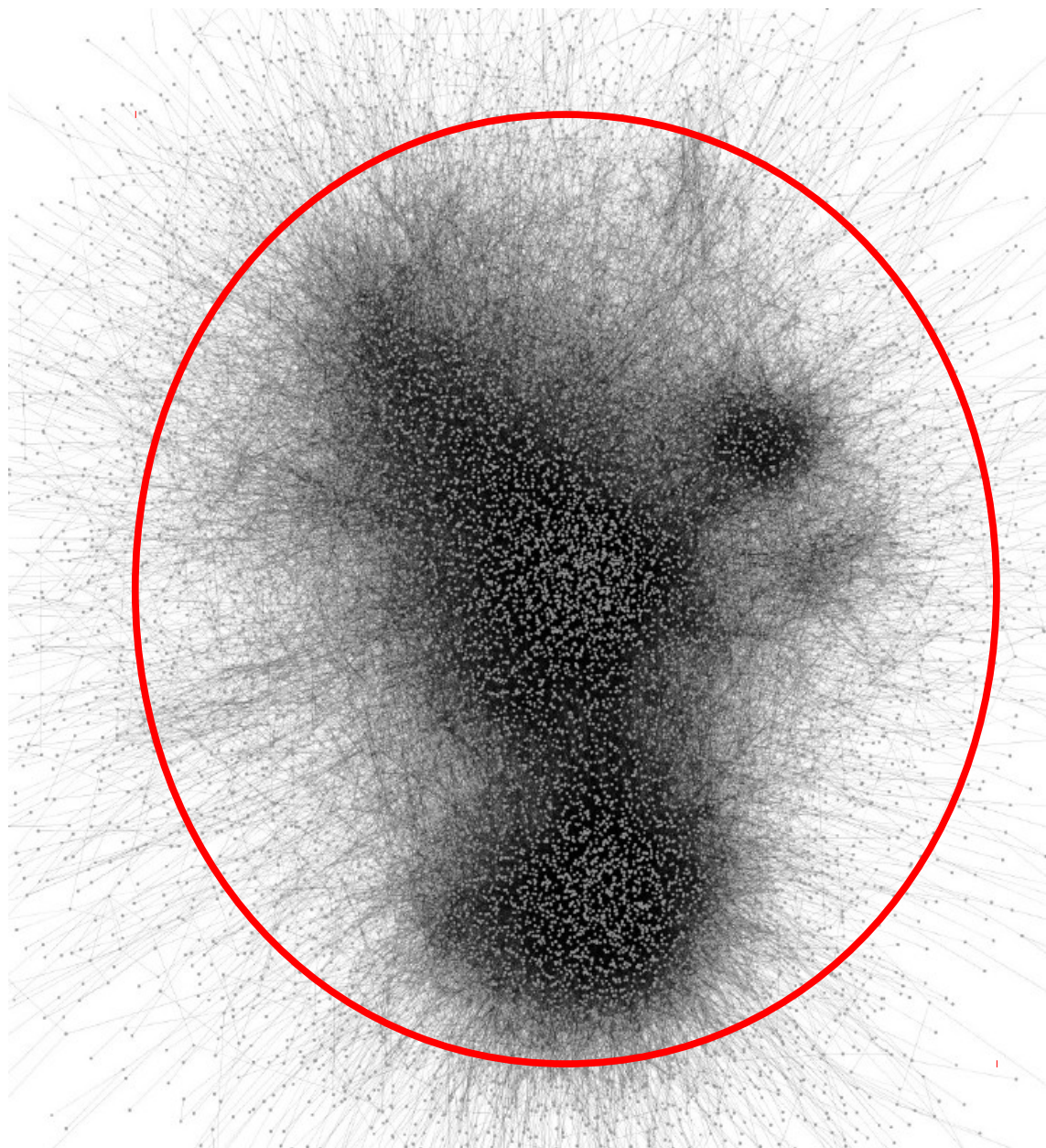
Communities

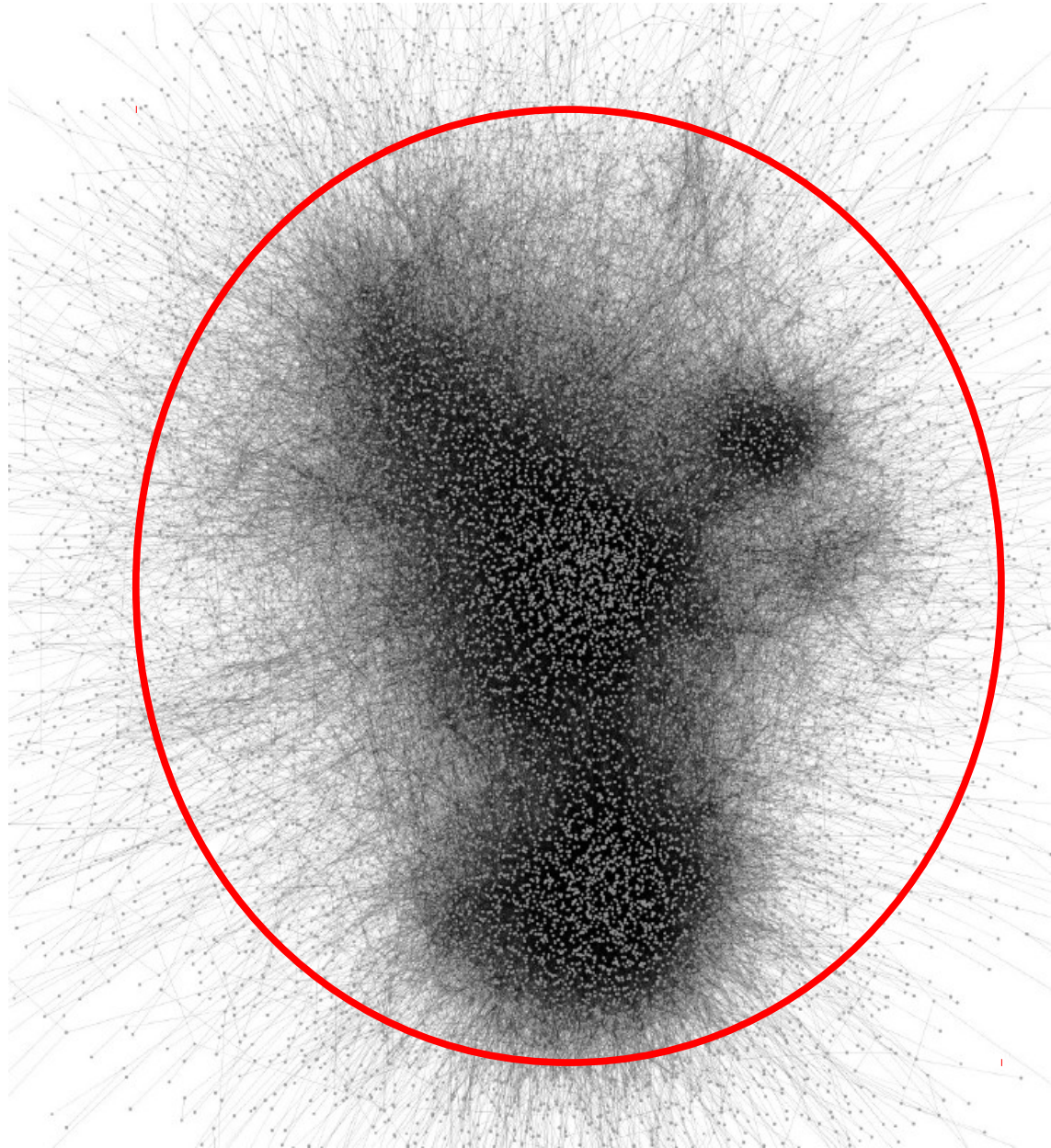
That's great!



Does the real world look like that?

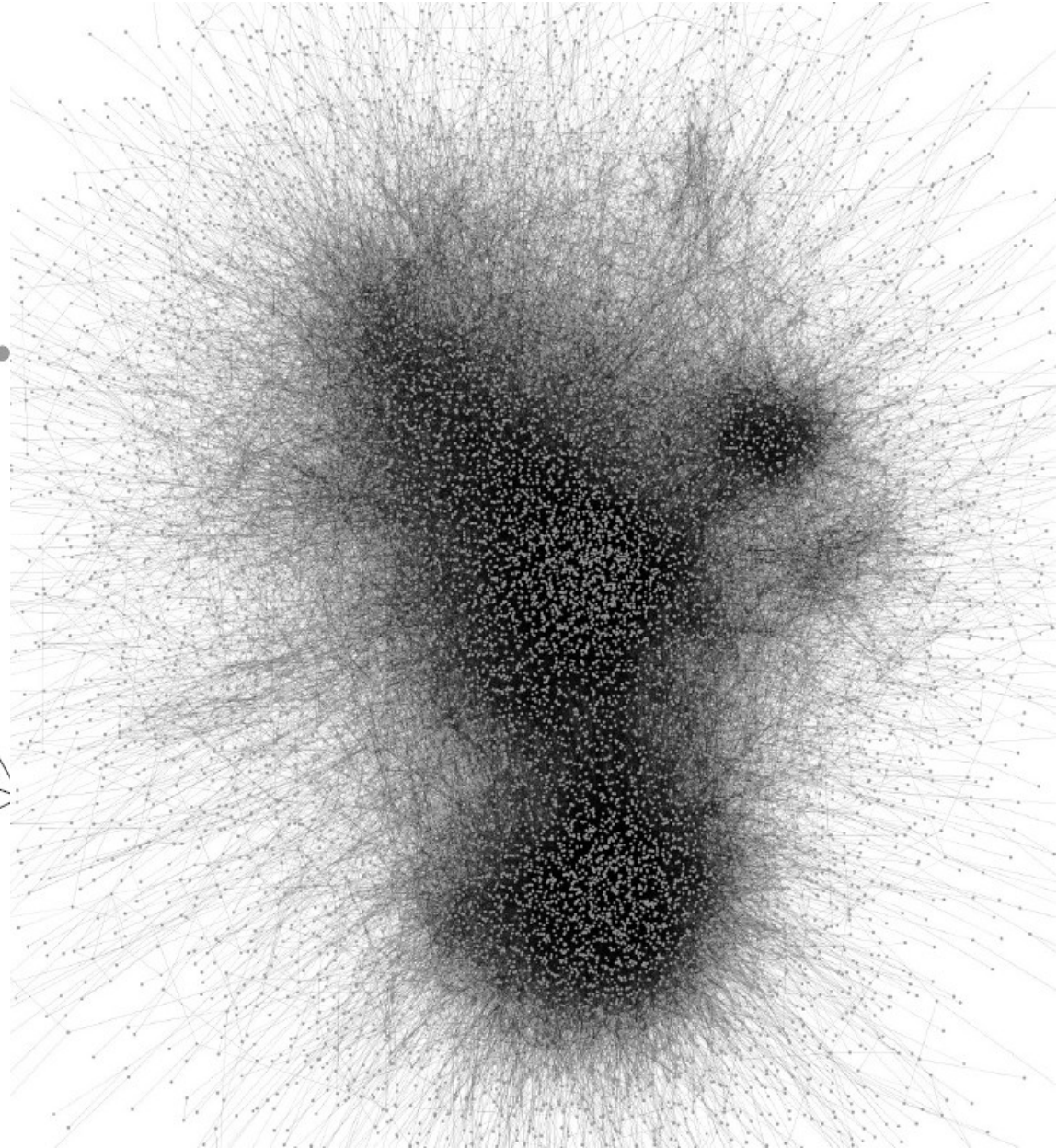
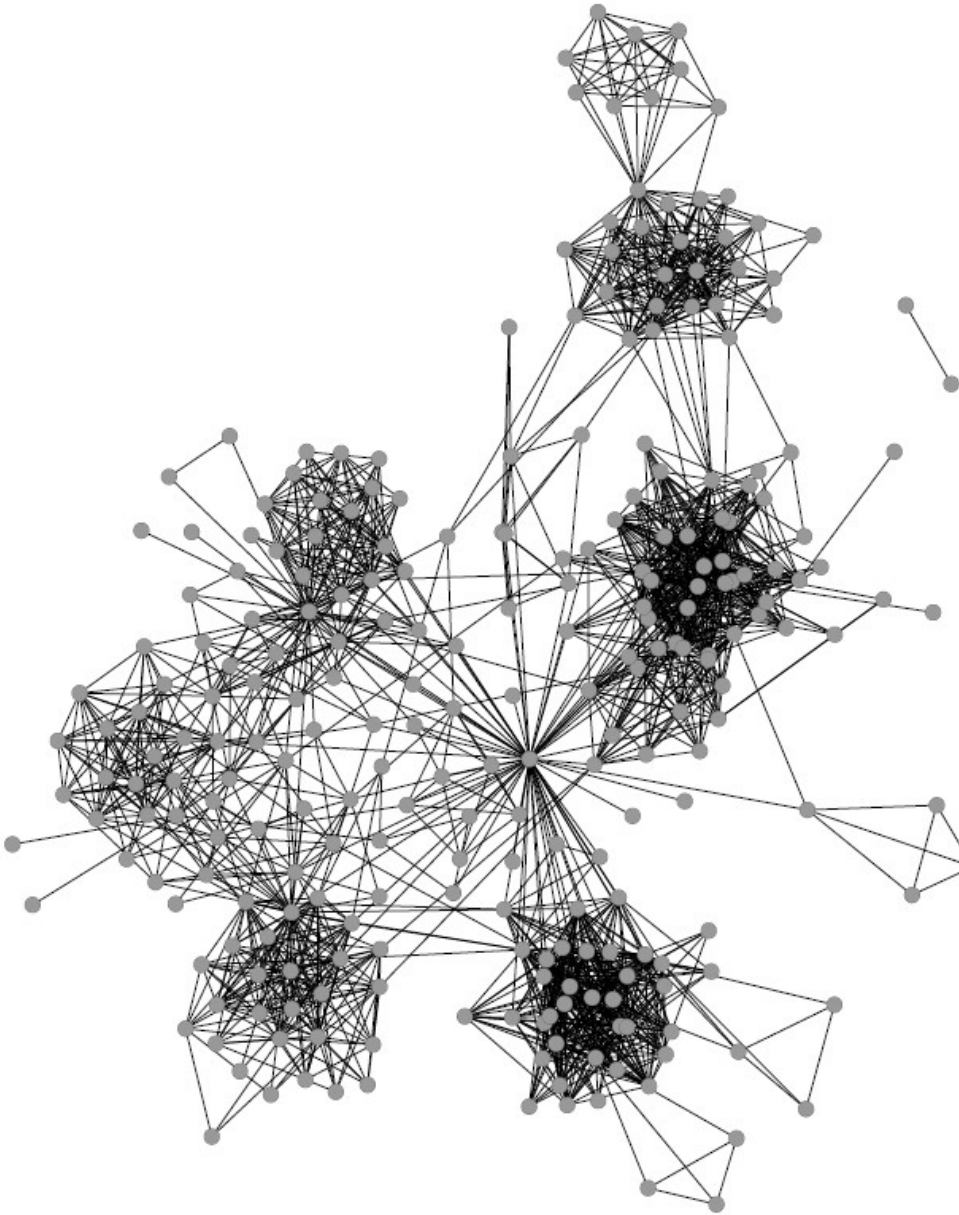




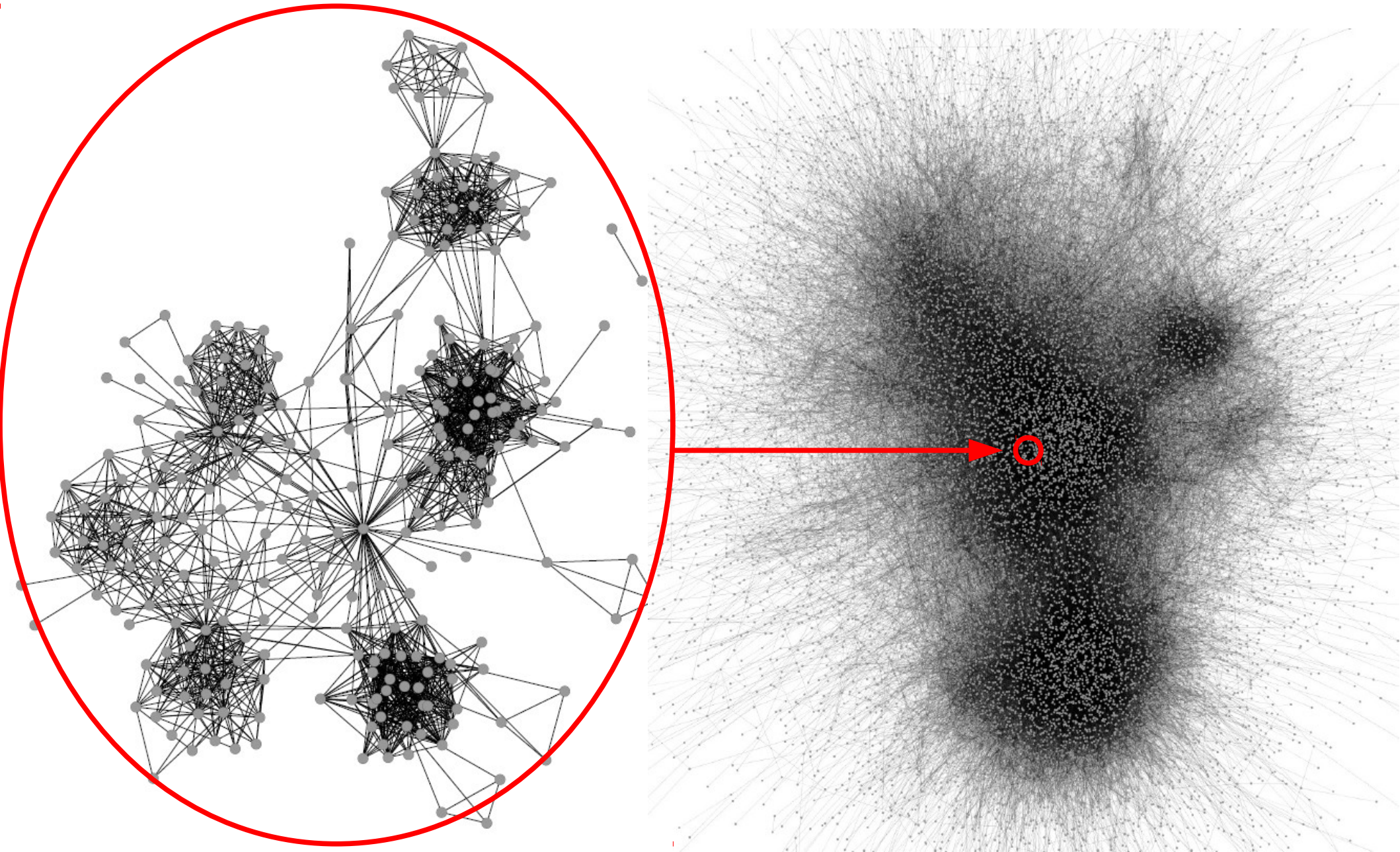


What the ... ?

Are they two different phenomena?

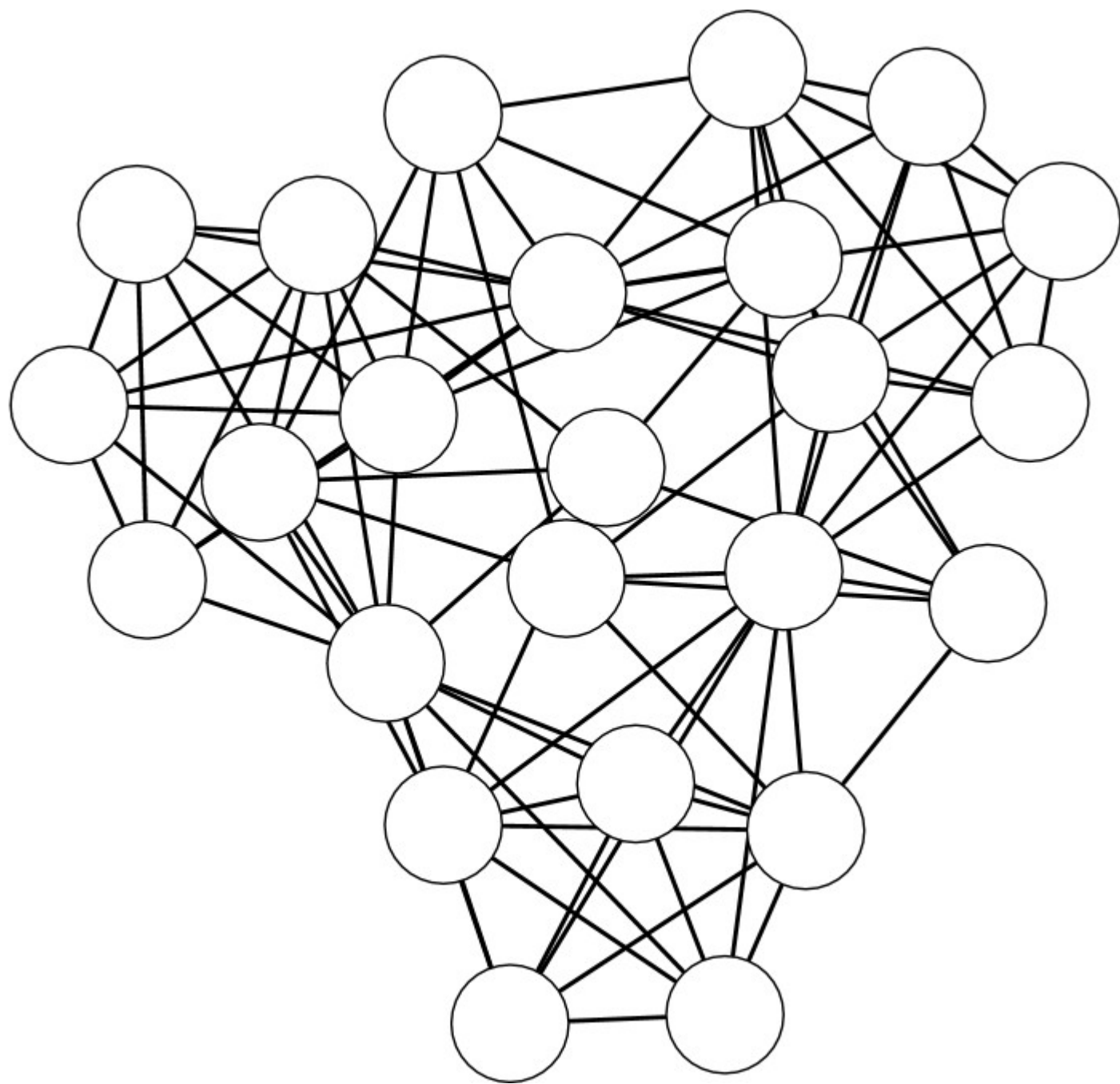


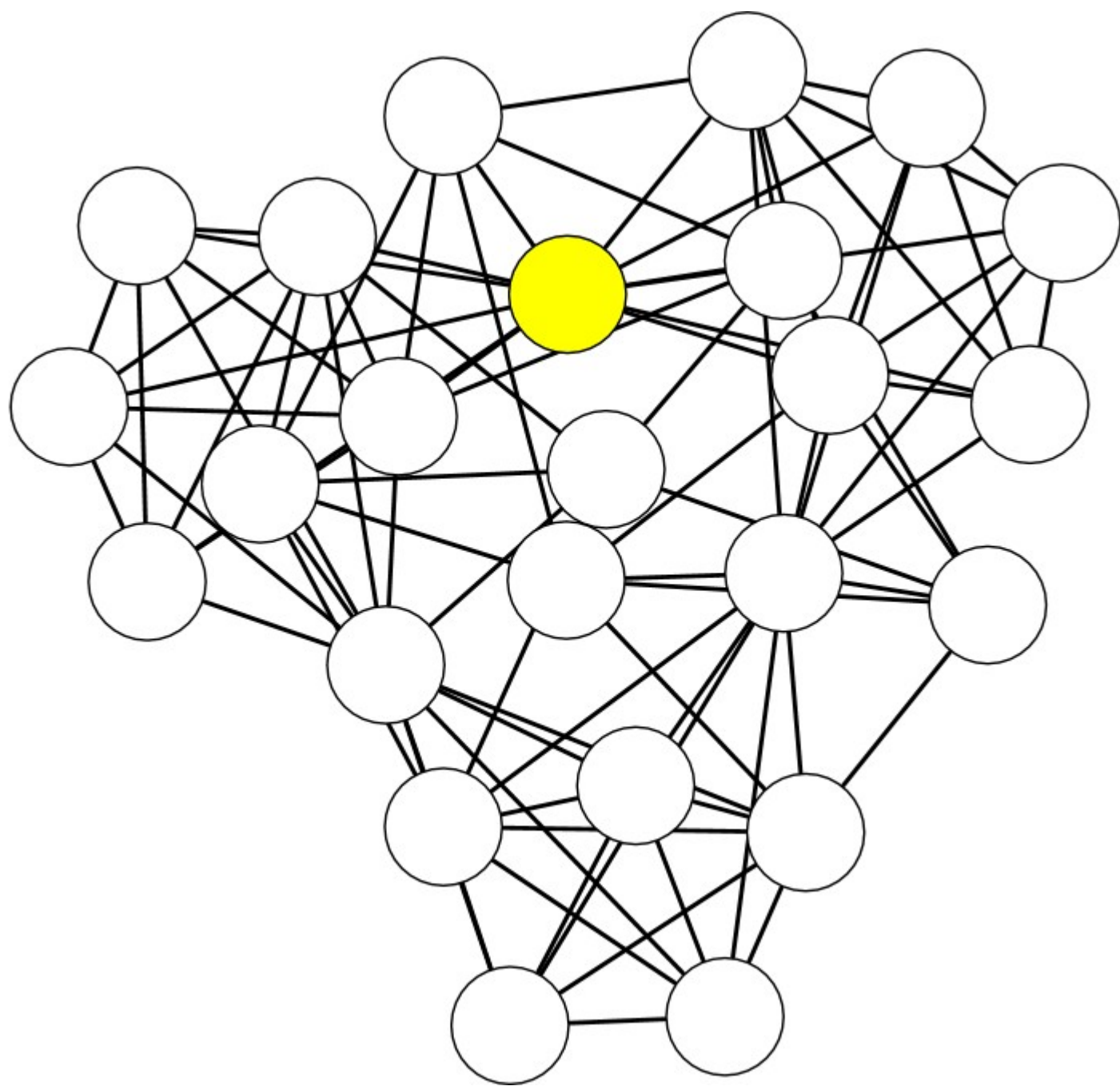
No!

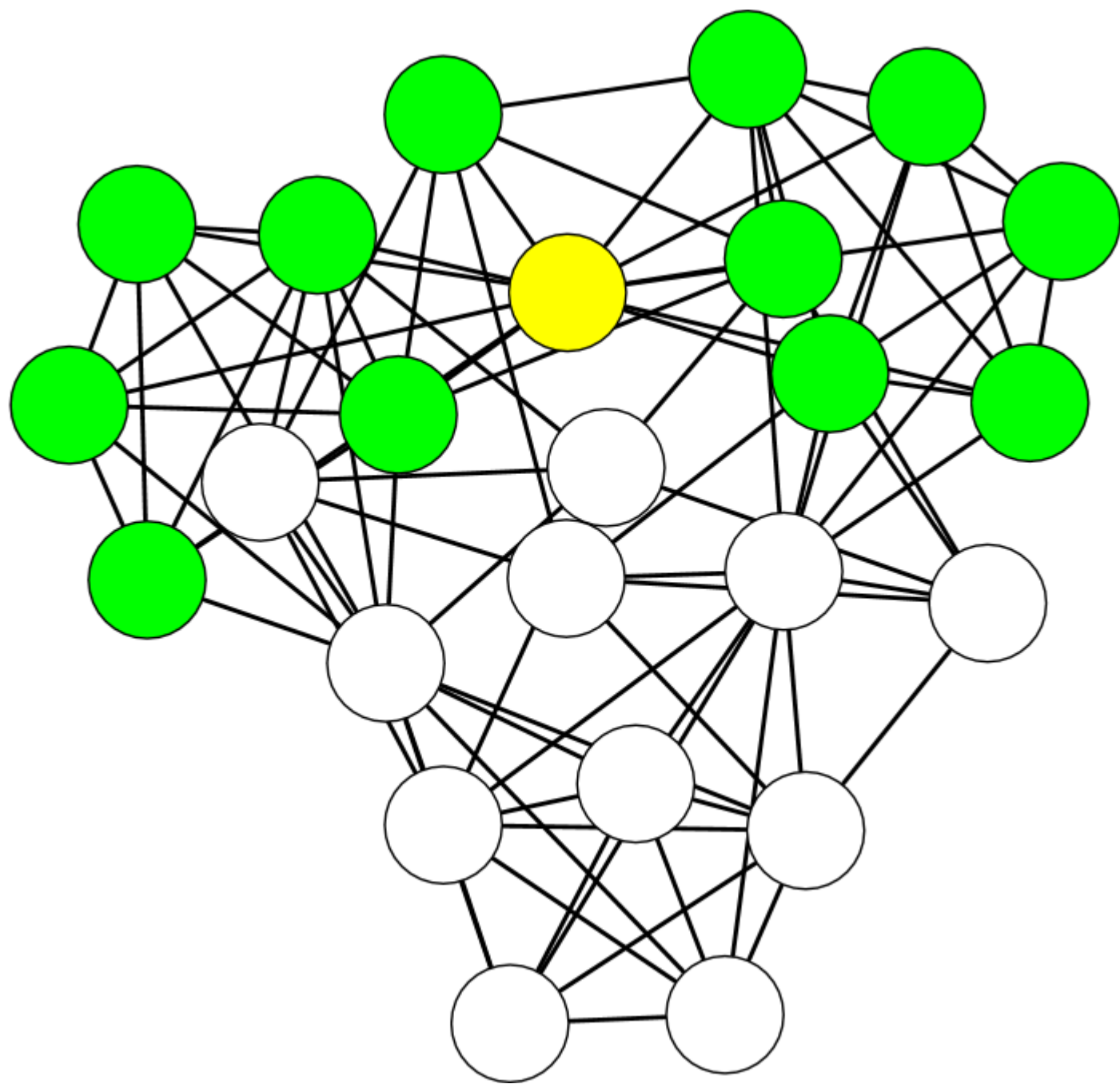


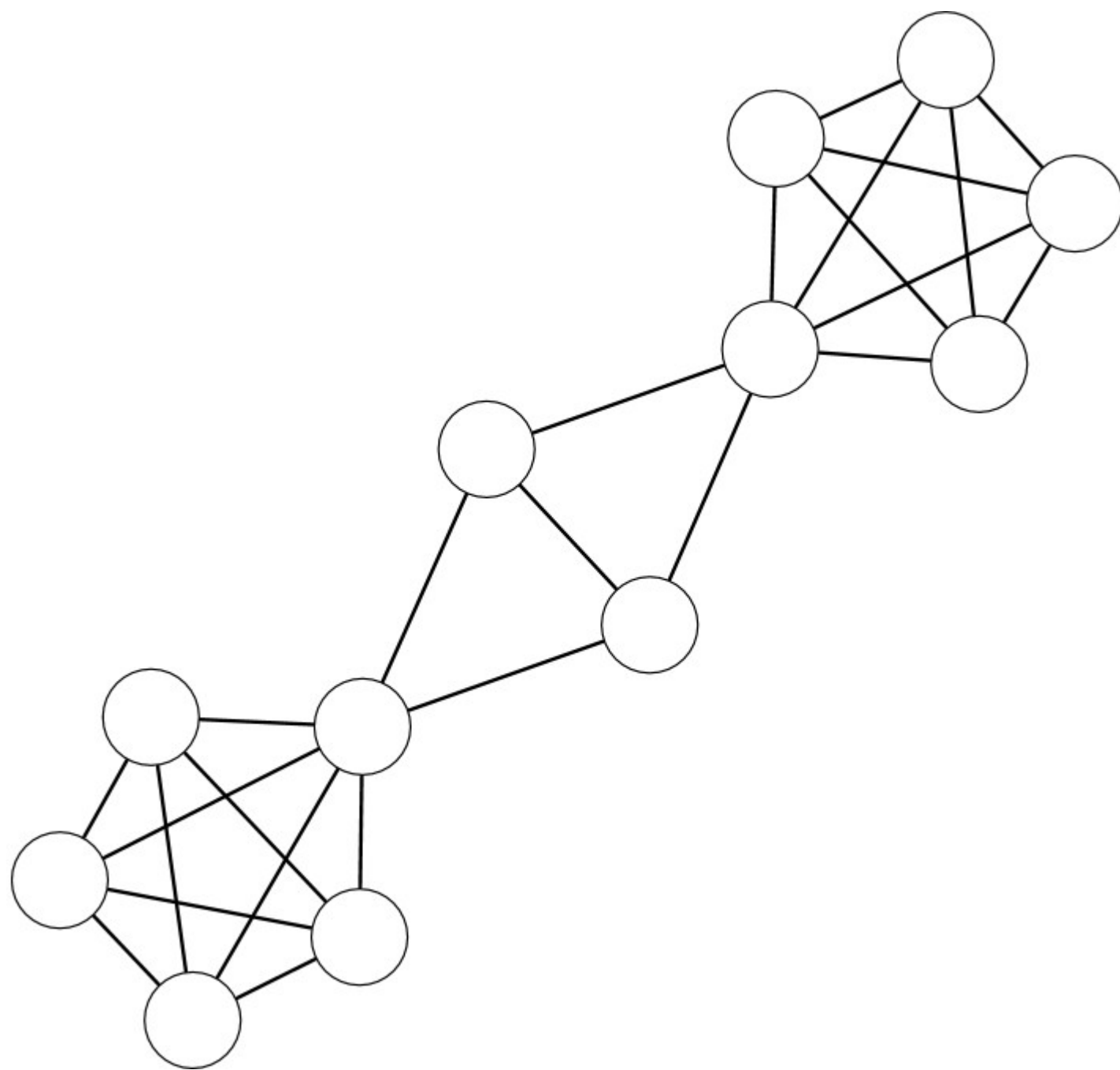
A Matter of Perspective

- The only difference is in the scale
- Locally, for each node the structure makes sense
- Globally, we are tangled in the complex overlap









DEMON Algorithm

- For each node n :
 - Extract the Ego Network of n
 - Remove n from the Ego Network
 - Perform a simple label propagation CD
 - Update the raw community set C
- For each row community c in C :

$$Merge(c, \mathcal{C}) = \begin{cases} \mathcal{C} & \exists c' \in \mathcal{C} : c \subseteq c' \\ \{c\} \cup \{c' \in \mathcal{C} \mid c' \not\subseteq c\} & \text{otherwise} \end{cases}$$

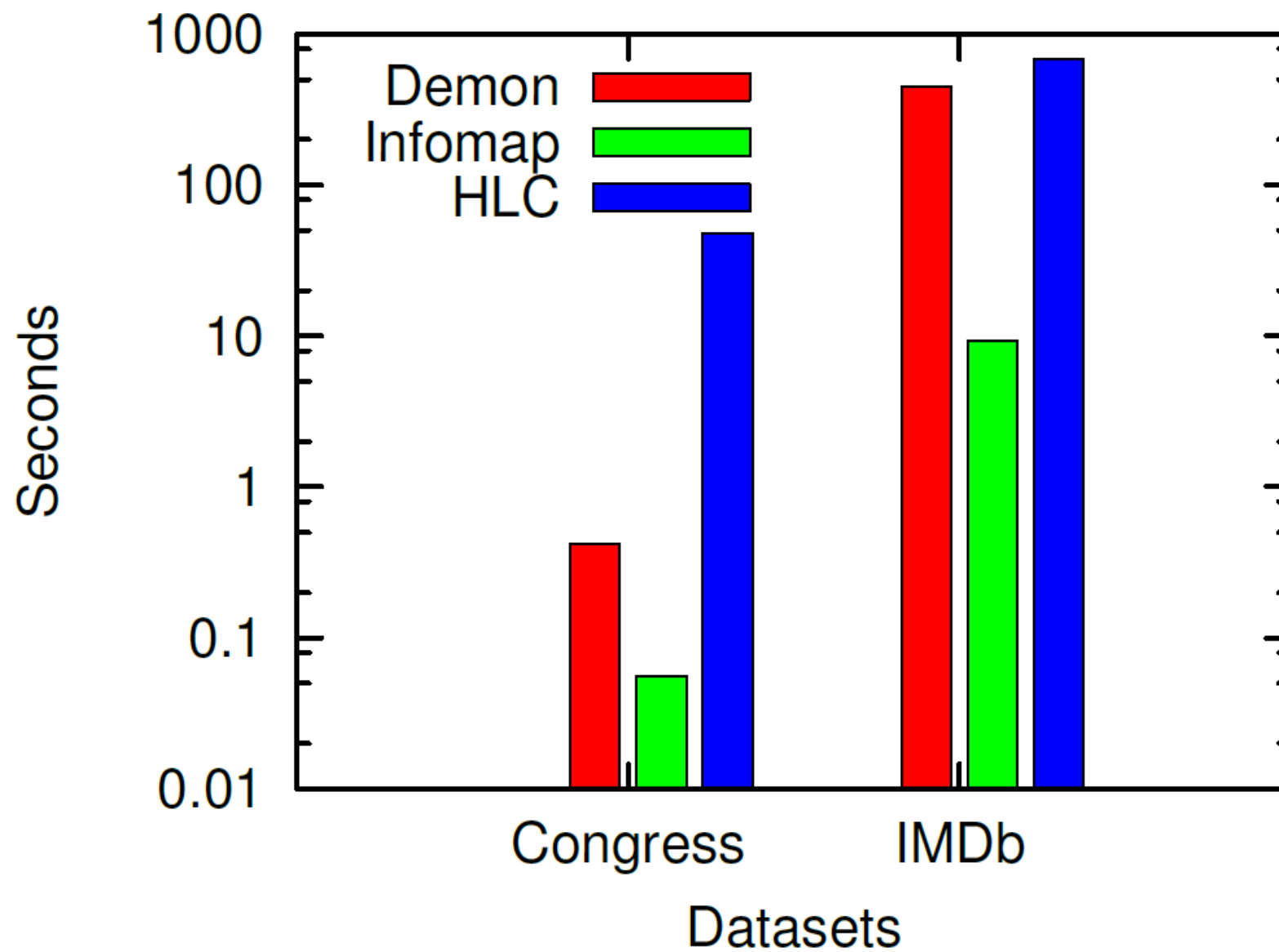
A nice property

$$DEMON(\mathcal{G} \cup \Delta\mathcal{G}, \mathcal{C}) = DEMON(\Delta\mathcal{G}, DEMON(\mathcal{G}, \mathcal{C}))$$

Experiments

- Networks:
 - Congress (nodes US politicians, connected if they co-sponsor the same bills)
 - IMDb (nodes Actors, connected if they play in the same movies)
- Compared Algorithms:
 - Infomap, non-overlapping state-of-the-art
 - Rosvall and Bergstrom “Maps of random walks on complex networks reveal community structure”, PNAS, 2008
 - HLC, overlapping state-of-the-art
 - Ahn, Bagrow and Lehmann “Link communities reveal multiscale complexity in networks”, Nature, 2010

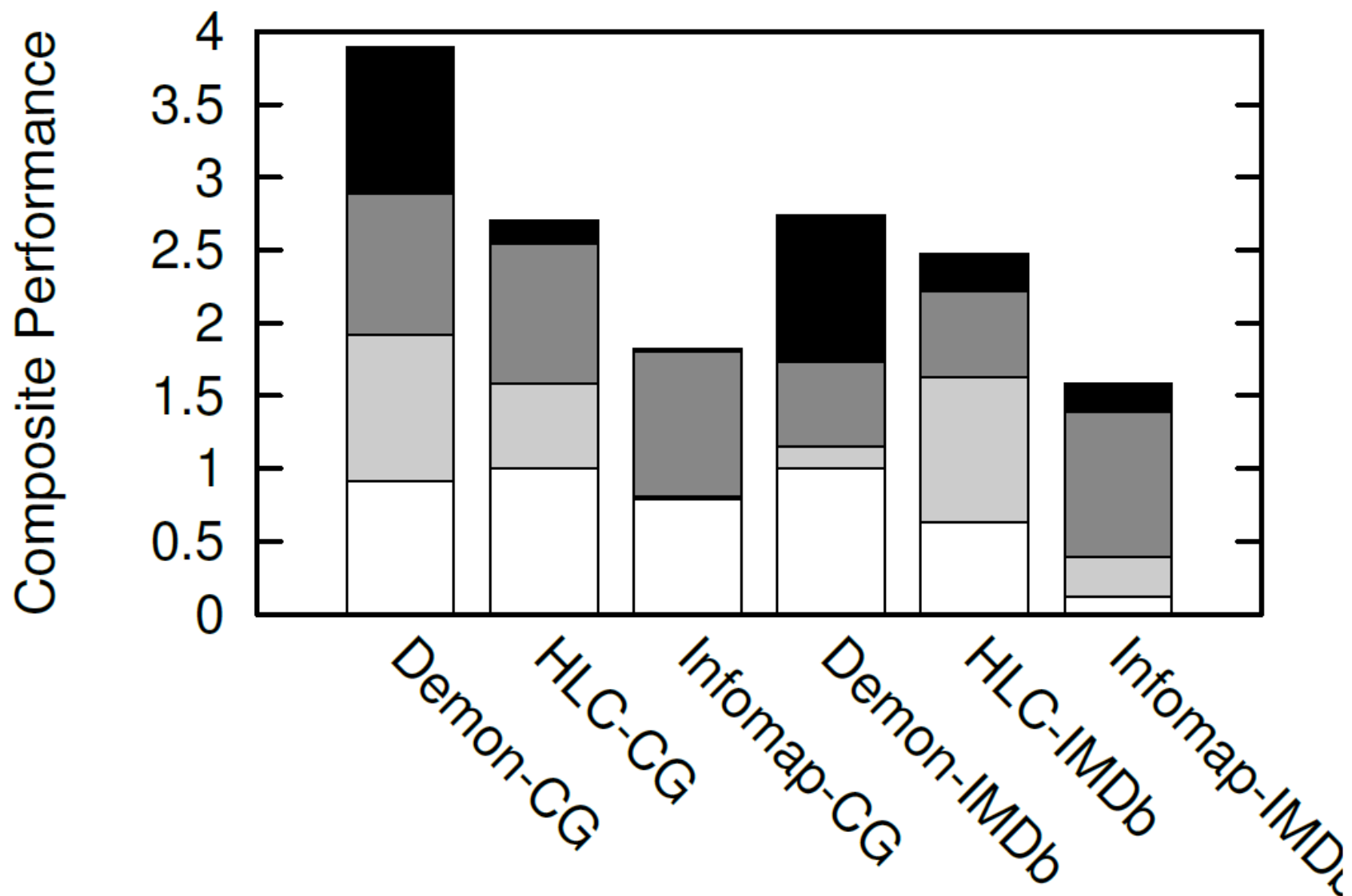
Runtimes



Evaluation

- **Community Quality:** Similar nodes share more qualitative attributes than dissimilar nodes;
- **Overlap Quality:** The more attributes, the more communities the node should be part of;
- **Community Coverage:** The proportion of nodes part of a non-trivial community;
- **Overlap Coverage:** How much overlap information the algorithm can return.

Evaluation



Conclusion

- A good intuition about the relationship between the global and the local structure
- Outperform over state-of-the-art
- A possible parallel implementation: very scalable

Thank you!

Questions?



coscia@di.unipi.it

