Group extraction for real-world networks

Lovro Šubelj[1], Neli Blagus & Marko Bajec

University of Ljubljana, Faculty of Computer and Information Science, Slovenia

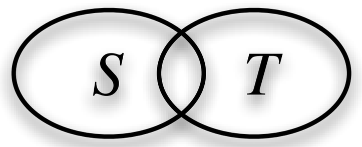
Background

Complex real-world networks contain characteristic groups of nodes with common linking pattern like densely linked **communities** [1]. These were the focus of most recent work and have diverse applications. However, many real-world networks also contain **other groups of nodes** that can be **overlapping** and other, whereas some parts of the networks reveal **no significant groups**.

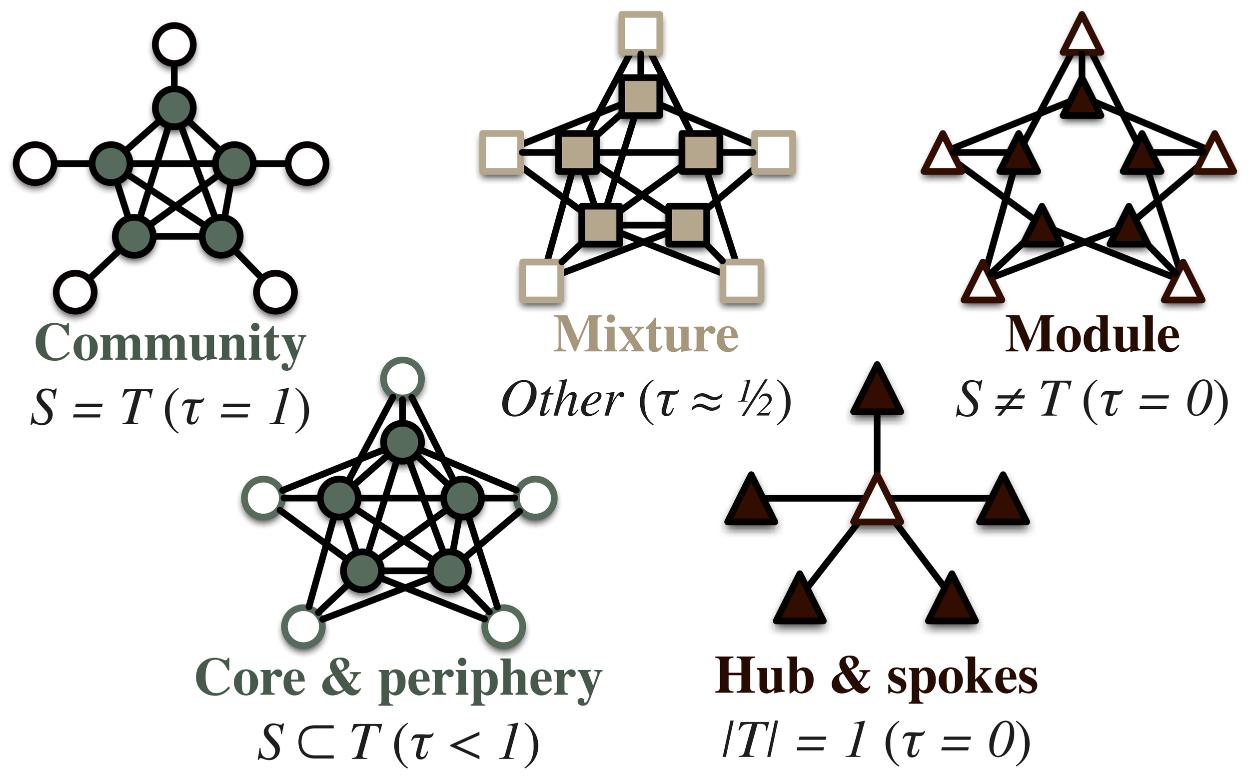
Group **formalism**

Let S be a **group of nodes**, T the **linking pattern** and τ the **group parameter**.

$$\tau(S,T) = \frac{|S \cap T|}{|S \cup T|}$$



Group **examples**



Group **criterion**

Let W be the **group criterion**, L the number of links and μ the (harmonic) mean size.

$$W(S,T) = \mu(S,T)\,(1-\mu(S,T))\left(\frac{L(S,T)}{|S||T|} - \frac{L(S,T^C)}{|S||T^C|}\right)$$

W is a **local asymmetric** criterion that **favors** the links between S and T, and **penalizes** for the links between S and T^C. (Note, however, that W **disregards** the links with both endpoints in S^C.) For $S = T$, W is consistent with a wide class of other models (e.g., *stochastic blockmodel*). [2]

Group **extraction**

A sequential extraction [2] of groups that can be **overlapping, nested etc.**

(1) **Find S and T** that optimize criterion W (e.g., *tabu search*).

(2) **Extract** only the **explained links** between S and T (and isolated nodes).

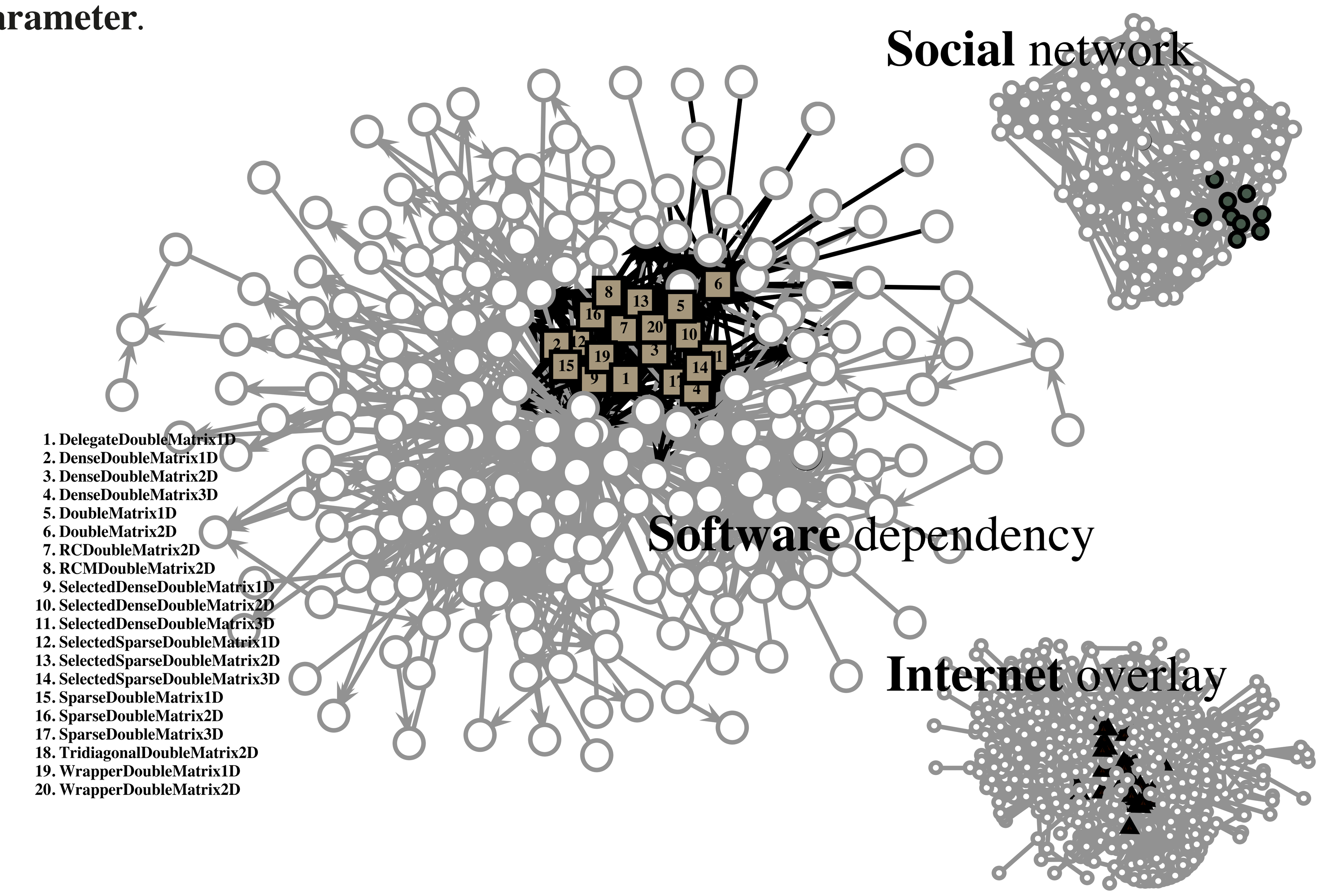
(3) **Repeat** until **W is larger than expected** in a random graph (by simulation).

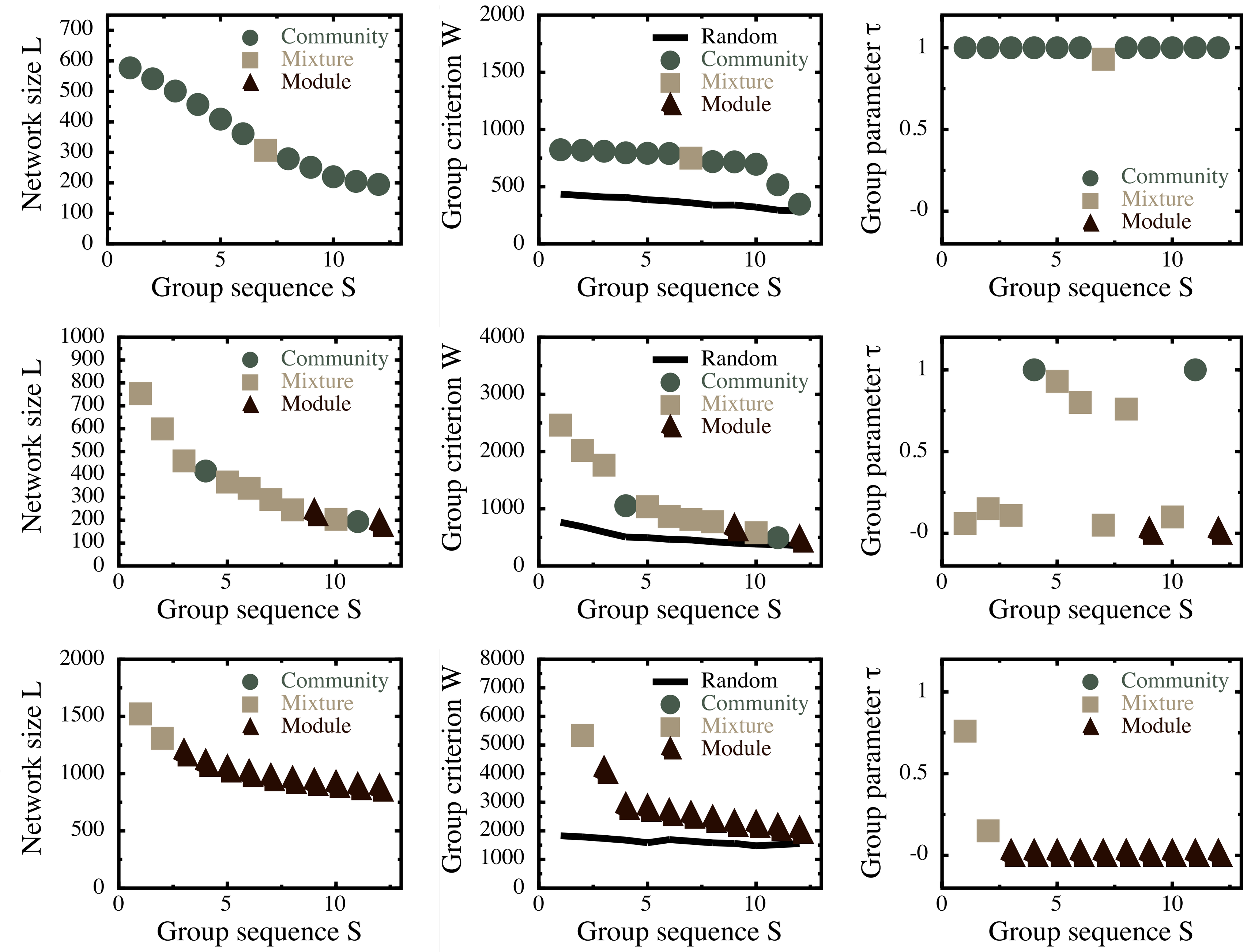
Contributions

1. A simple **formalism and criterion for general groups** of nodes.
2. An **adequate extraction procedure** for statistically significant groups.
3. **Characterization of the group structure** of different real-world networks.

What are characteristic groups of nodes in real-world networks? **Network (type) dependent.**
What portion of network links is explained by the group structure? **Between 60% and 90%.**
What portion of network nodes is included in the group structure? **More than 50%.**

Groups in **real-world networks**





Network	Nodes	Links	Group #	Group $\lvert S\rvert$	Group τ	Community	Core	Mixture	Module	Background
						% Links (% nodes)				
Author collaborat. [3]	1589	2742	160	5.6	**0.94**	**71% (47%)**	0% (0%)	6% (5%)	1% (1%)	*22% (47%)*
American football [1]	115	613	13	8.6	**0.88**	**59% (83%)**	9% (11%)	3% (7%)	0% (0%)	*29% (98%)*
Lucene search engine	1657	6808	123	12.1	**0.55**	19% (25%)	1% (2%)	**30% (24%)**	**38% (34%)**	*11% (49%)*
Colt computing [4]	227	963	15	10.3	**0.41**	7% (11%)	5% (6%)	**69% (49%)**	4% (6%)	*15% (64%)*
Word adjacency [3]	112	425	4	11.2	**0.28**	0% (0%)	0% (0%)	**34% (33%)**	25% (15%)	*41% (99%)*
Internet overlay [5]	767	1857	33	10.6	**0.08**	0% (1%)	12% (4%)	13% (7%)	**34% (35%)**	*41% (80%)*
Southern women [6]	32	89	2	4.3	**0.00**	0% (0%)	0% (0%)	0% (0%)	**80% (41%)**	*20% (47%)*

All extracted groups are statistically significant at *1%* level.

References

[1] Girvan, M. & Newman, M.E.J.: Community structure in social and biological networks. **P. Natl. Acad. Sci. USA** 99(12), 7821–7826 (2002).
[2] Zhao, Y., Levina, E., & Zhu, J.: Community extraction for social networks. **P. Natl. Acad. Sci. USA** 108(18), 7321–7326 (2011).
[3] Newman, M. E. J.: Finding community structure in networks using the eigenvectors of matrices. **Phys. Rev. E** 74(3), 036104 (2006).
[4] Šubelj, L. & Bajec, M.: Community structure of complex software systems: Analysis and applications. **Physica A** 390(16), 2968-2975 (2011).
[5] Leskovec, J., Kleinberg, & J., Faloutsos, C.: Graphs over time: Densification laws, shrinking diameters and possible explanations. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Chicago, IL, USA, 2005), pp. 177–187.
[6] Davis, A., Gardner, B.B., & Gardner, M.R.: *Deep South* (Chicago University Press, Chicago, 1941).

[1]**Corresponding author:** `lovro.subelj@fri.uni-lj.si`