

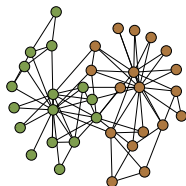
NETWORK COMMUNITY DETECTION IN PRACTICAL SCENARIOS

Lovro Šubelj

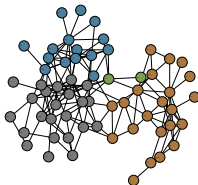
University of Ljubljana
Faculty of Computer and Information Science

ARS '15

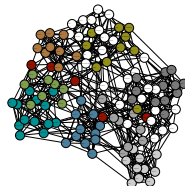
NETWORK COMMUNITY STRUCTURE



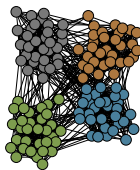
Karate club



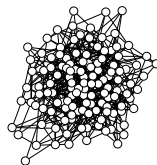
Bottlenose dolphins



American football



Synthetic graph



Random graph

Communities are cohesive subgroups of sparse networks.

NETWORK COMMUNITY DETECTION

- graph partitioning,
- hierarchical clustering,
- modularity optimization,
- statistical inference,
- spectral methods,
- map equation,
- dynamics etc.

Girvan, M. & Newman, M. E. J., *P. Natl. Acad. Sci. USA* **99**, 7821–7826 (2002).

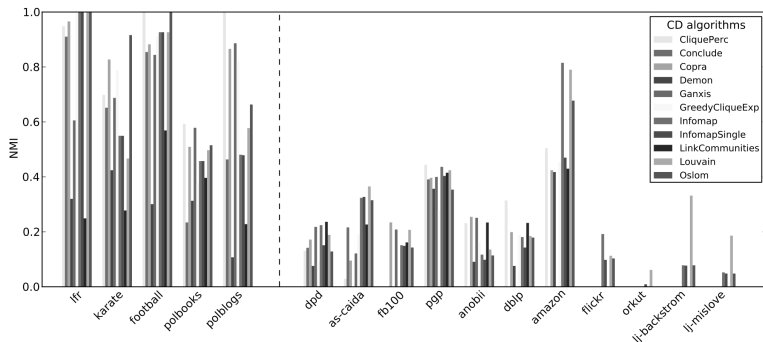
Fortunato, S., *Phys. Rep.* **486**, 75–174 (2010).

LABEL PROPAGATION ALGORITHM

Raghavan, U. N., Albert, R. & Kumara, S., *Phys. Rev. E* **76**, 036106 (2007).

Šubelj, L. & Bajec, M., *Phys. Rev. E* **83**, 036103 (2011) etc.

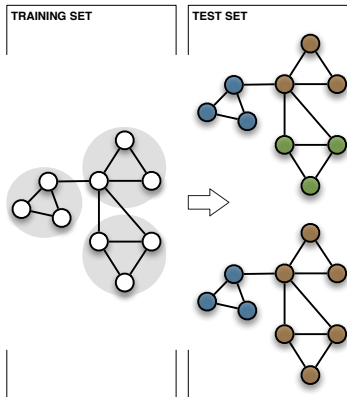
LARGE-SCALE COMMUNITY DETECTION



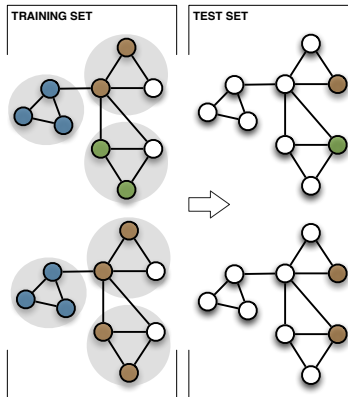
Hric, D., Darst, R. K. & Fortunato, S., *Phys. Rev. E* **90**, 062805 (2014).

COMMUNITY DETECTION TASKS

EXPLORATORY TASK



PREDICTIVE TASK

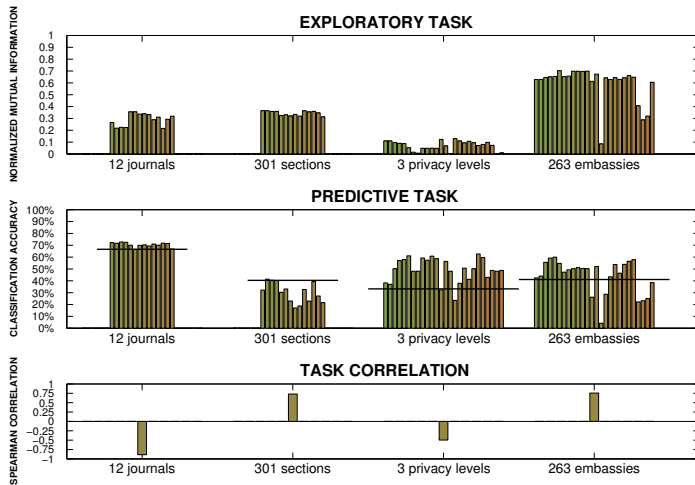


APS & WIKILEAKS NETWORKS

DATA	APS 1893-2013	WikiLeaks 1966-2010
NETWORK	citation 526,527 papers 5,989,263 citations	reference 52,416 cables 78,506 references
CLUSTERS	12 journals 301 sections	3 privacy levels 263 embassies
SETTING	14 algorithms	26 algorithms
TRAINING	1893-2012	1966-2009
TEST	2013 (4%)	2010 (17%)

Non-overlapping and cohesive ground truth clusters.

APS & WIKILEAKS RESULTS

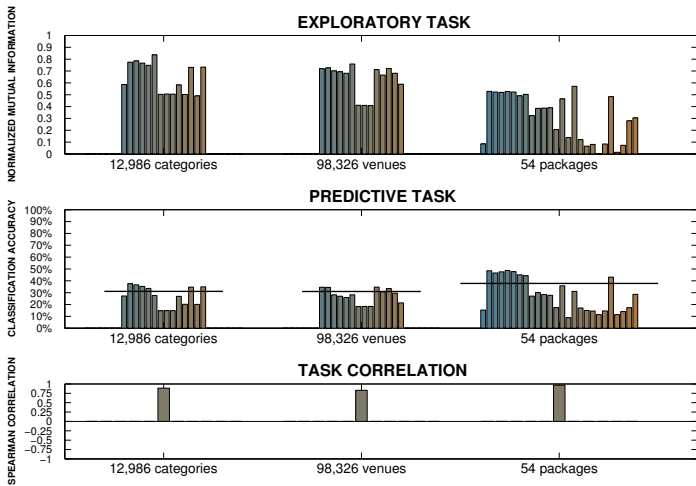


YOUTUBE, DBLP & JAVA NETWORKS

DATA	YouTube	DBLP	java
NETWORK	social 39,841 users 224,235 friends.	collaboration 317,080 authors 1,049,866 collabs.	software 2,378 classes 14,619 depends.
CLUSTERS	12,986 groups	98,326 venues	54 packages
SETTING	14 algorithms	14 algorithms	26 algorithms
TRAINING	leave-one-out	leave-one-out	leave-one-out

Overlapping or non-cohesive ground truth clusters.

YOUTUBE, DBLP & JAVA RESULTS



COMMUNITY DETECTION IN PRACTICE

Take-home message:

- community information **is** useful in practice,
- for lots of clusters, same algorithms for both tasks,
- for few clusters, different algorithms for different tasks.

Future work:

- beyond majority classification,
- overlapping and non-cohesive clusters,
- descriptive, inferential, causal and mechanistic tasks.

LOVRO ŠUBELJ

lovro.subelj@fri.uni-lj.si

<http://lovro.lpt.fri.uni-lj.si>