# network *mining*

introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2020/21

# mining *overview*

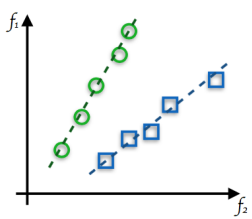
how to *mine* different *graphs/networks*?

how to *mine* network *nodes/links*?

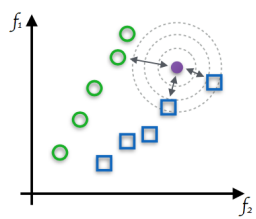
*node/link clustering* → revealing *similarity clusters*

*node/link classification* → predicting *discrete labels*

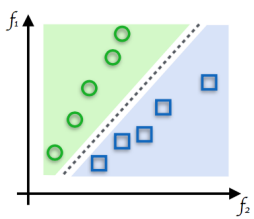
*node/link regression* → predicting *numerical values*



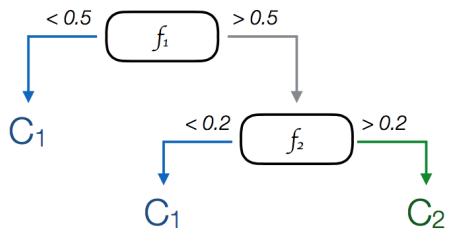

Regression



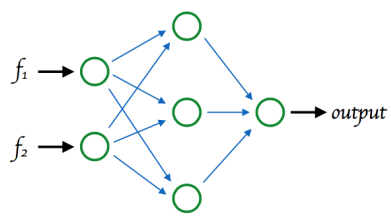

kNN




SVM




Decision Tree




ANN

# mining *clustering*

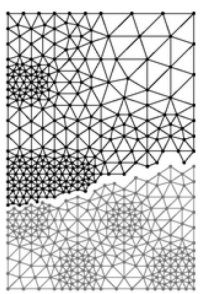
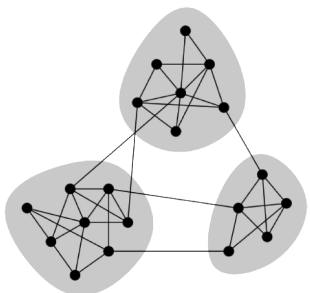
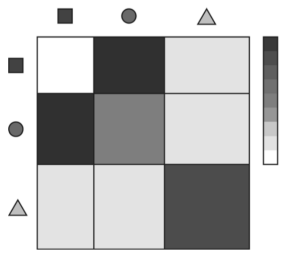
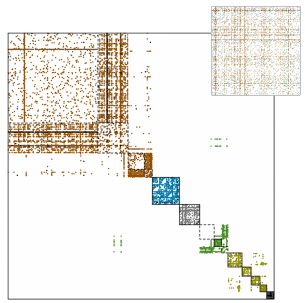
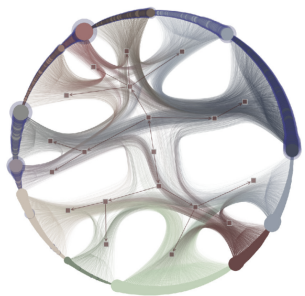
how to *cluster* network *nodes/links*?

— *graph partitioning* and *community detection* methods

for connected *assortative clusters* based on homophily

— *(stochastic) blockmodeling* and *role discovery* methods

also disconnected *disassortative clusters* based on equivalence

for *survey/user guide* see [For10, FH16]

# mining *classification*

how to *classify* network *nodes/links*?

— *relational learning* and *link mining* methods

*machine learning* methods using network structure

— *node/link structure* used as *classification features*

centrality, bridging, fragments, *egonets*, *clusters* etc.

node features from local *random walk exploration* →

*DeepWalk* [PARS14]
only *homophily*

*node2vec* [GL16]
also *equivalence*

*struc2vec* [FRS17]
only *structure*

for *survey* see [BCM11, ZPS+16]

# mining *classification*

— *classification by clustering* in *APS citation* network [Šub15b]

| | | 7 **journals** | | 91 **sections** | |
|---|---|---|---|---|---|
| **class** | **method** | **NMI** | **CA** | **NMI** | **CA** |
| clusters | spectrum | 0.361 | 59.8% | 0.380 | 38.6% |
| | *modularity* | 0.339 | 68.1% | 0.426 | 37.1% |
| | *map equation* | 0.232 | 71.3% | 0.416 | 48.1% |
| | *block model* | 0.243 | 69.6% | 0.392 | 45.3% |
| baselines | *neighbors* | - | 63.9% | - | 46.5% |
| | *2-neighbors* | - | 71.5% | - | 50.4% |
| | network | - | 27.6% | - | 17.9% |

— *. . . by clustering* in *WikiLeaks reference* network [Šub15a]

| | | 3 **privacies** | | 246 **embassies** | |
|---|---|---|---|---|---|
| **class** | **method** | **NMI** | **CA** | **NMI** | **CA** |
| clusters | *spectrum* | 0.003 | 49.1% | 0.658 | 47.9% |
| | *modularity* | 0.048 | 59.2% | 0.699 | 52.3% |
| | map equation | 0.088 | 33.1% | 0.654 | 37.1% |
| | *block model* | 0.035 | 56.5% | 0.625 | 37.6% |
| baselines | neighbors | - | 14.2% | - | 15.0% |
| | 2-neighbors | - | 27.7% | - | 31.6% |
| | network | - | 49.1% | - | 1.4% |

# mining *regression*

how to *regress*/*rank* network *nodes*/*links*?

— *relational learning* and *link mining* methods

*machine learning* methods using network structure

— *node*/*link structure* used as *regression features*

*centrality*, *bridging*, fragments, egonets, clusters etc.

node features from local *random walk exploration* →

*DeepWalk* [PARS14]
only *homophily*

*node2vec* [GL16]
also *equivalence*

*struc2vec* [FRS17]
only *structure*

for *survey* see [BCM11, ZPS+16]

# mining *regression*

— *regression by centrality* in *Slovenian highways* network [HW1]

| | | traffic loads | |
|---|---|---|---|
| **class** | **centrality** | **Pearson** | **Spearman** |
| spectral | degree | 0.278 | 0.275 |
| | eigenvector | 0.010 | 0.241 |
| | PageRank | 0.121 | −0.149 |
| distance | *closeness* | 0.623 | 0.634 |
| | *betweenness* | 0.628 | 0.651 |
| clustering | standard | 0.000 | 0.000 |
| | $\mu$-corrected | 0.000 | 0.230 |

— *. . . by centrality* in *Wikipedia web of trust* network [MAC11]

| | | users trust | |
|---|---|---|---|
| **class** | **centrality** | **Pearson** | **Spearman** |
| spectral | *degree* | 0.013 | 0.316 |
| | eigenvector | 0.026 | 0.118 |
| | PageRank | 0.001 | 0.182 |
| clustering | standard | 0.134 | 0.115 |
| | *$\mu$-corrected* | 0.127 | 0.199 |

# mining *references*


A.-L. Barabási.
*Network Science*.
Cambridge University Press, Cambridge, 2016.

Smriti Bhagat, Graham Cormode, and S. Muthukrishnan.
Node classification in social networks.
In *Social Network Data Analytics*. Springer, 2011.

Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj.
*Exploratory Social Network Analysis with Pajek: Expanded and Revised Second Edition*.
Cambridge University Press, Cambridge, 2011.

David Easley and Jon Kleinberg.
*Networks, Crowds, and Markets: Reasoning About a Highly Connected World*.
Cambridge University Press, Cambridge, 2010.

Ernesto Estrada and Philip A. Knight.
*A First Course in Network Theory*.
Oxford University Press, 2015.

Santo Fortunato and Darko Hric.
Community detection in networks: A user guide.
*Phys. Rep.*, 659:1–44, 2016.

Santo Fortunato.
Community detection in graphs.
*Phys. Rep.*, 486(3-5):75–174, 2010.

# mining *references*


Daniel R. Figueiredo, Leonardo F. R. Ribeiro, and Pedro H. P. Saverese.

struc2vec: Learning node representations from structural identity.
In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1–9, 2017.

Aditya Grover and Jure Leskovec.

node2vec: Scalable feature learning for networks.
In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 855–864, 2016.

Silviu Maniu, Talel Abdessalem, and Bogdan Cautis.

Casting a web of trust over Wikipedia: An Interaction-based approach.
In *Proceedings of the International Conference on World Wide Web*, pages 87–88, New York, NY, USA, 2011.

Mark E. J. Newman.

*Networks*.
Oxford University Press, Oxford, 2nd edition edition, 2018.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena.

DeepWalk: Online learning of social representations.
In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 701–710, 2014.

Lovro Šubelj.

Exploratory and predictive tasks of network community detection.
In *Proceedings of the International Conference on Network Science*, page 1, Zaragoza, Spain, 2015.

Lovro Šubelj.

Large network community detection in practical scenarios.
In *Proceedings of the International Workshop on Social Network Analysis*, page 78, Capri, Italy, 2015.

# mining *references*

M. Zanin, D. Papo, P. A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik, and S. Boccaletti.
Combining complex networks and data mining: Why and how.
*Phys. Rep.*, 635:1–44, 2016.