

label propagation for clustering

“razvrščanje vozlišč omrežja z izmenjavo oznak”

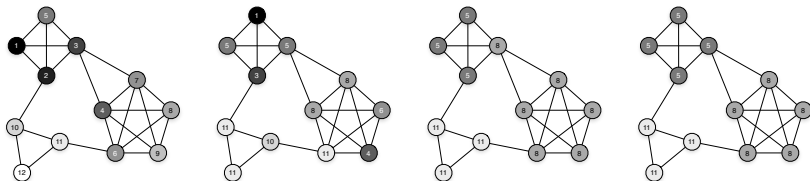
Lovro Šubelj
University of Ljubljana
Faculty of Computer and
Information Science

NetSlo '20

label propagation **animation**

clusters of nodes represented by **colors**

label propagation method



(**setting**) **cluster** of node i represented by its **label** c_i

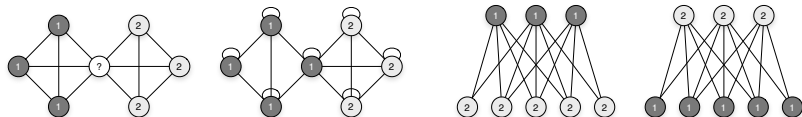
(**initialization**) put each node i in **own cluster** e.g. $c_i = i$

(**propagation**) label c_i set to **most frequent** in neighborhood Γ_i

$$c_i = \operatorname{argmax}_c |\{j \in \Gamma_i : c_j = c\}|$$

(**convergence**) propagate **until no** node i **changes** its label c_i

label propagation details



(**ties**) label ties resolved **randomly with retention**

(**order**) labels propagated **asynchronously** for convergence

(**links**) generalization to **weighted multigraphs** with adjacency A

$$c_i = \operatorname{argmax}_c |\{j \in \Gamma_i : c_j = c\}| = \operatorname{argmax}_c \sum_j A_{ij} \delta(c_j, c)$$

(**equilibrium**) propagate until convergence followed **by floodfill**

label propagation algorithm

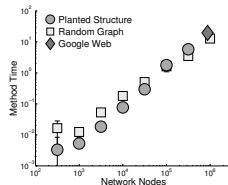
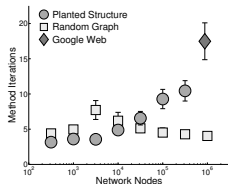
(**optimization**) label propagation is equivalent to **Potts model**

$$\mathcal{F}(\{c\}) = \sum_{ij} A_{ij} \delta(c_i, c_j)$$

(**optimum**) revealed structure is local & **not global optimum**

(**time**) complexity **almost linear** $\mathcal{O}(m^{1.2})$ in number of links m

(**terminology**) ... = **relocation algorithm** = local greedy optim.



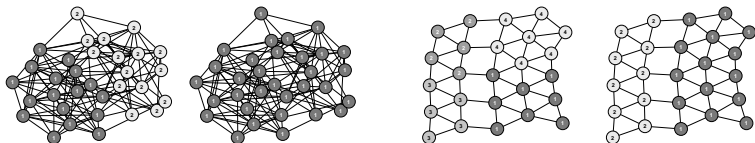
label propagation **advances**

(**modularity**) constrained label propagation is **Louvain algorithm**

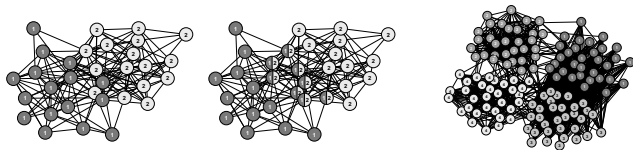
$$c_i = \operatorname{argmax}_c \sum_j \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_j, c)$$

(**preferences**) label propagation **using (anti)position** p_i of node i

$$c_i = \operatorname{argmax}_c \sum_j p_j A_{ij} \delta(c_j, c) \quad c_i = \operatorname{argmax}_c \sum_j (1 - p_j) A_{ij} \delta(c_j, c)$$



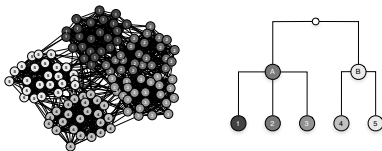
label propagation clusters



(left) communities, (middle) overlaps & (right) equivalences

$$c_i = \operatorname{argmax}_c \sum_j A_{ij} \delta(c_j, c) + \sum_{kj \neq i} \frac{1}{k_k - 1} A_{ik} A_{kj} \delta(c_j, c)$$

(bottom) hierarchical detection of nested clusters' dendrogram

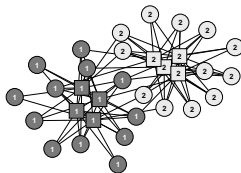


label propagation **networks**

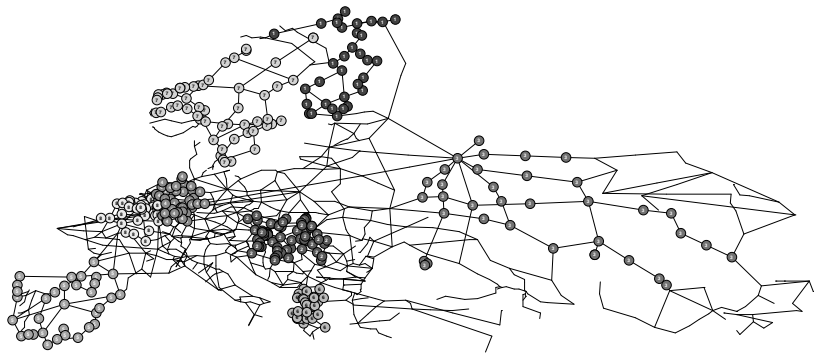
- (**weights**) **straightforward** for weighted multigraphs \uparrow
- (**directions**) there is **no general** method for directed graphs
- (**signs**) signed graphs with m_p **positive** & m_n **negative** links

$$c_i = \operatorname{argmax}_c \sum_j \left\{ \begin{array}{ll} 1/m_p & \text{for } A_{ij} \geq 0 \\ 1/m_n & \text{for } A_{ij} < 0 \end{array} \right\} A_{ij} \delta(c_j, c)$$

- (**multipartite**) labels propagated **synchronously** in each partition

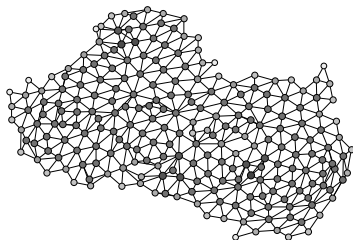
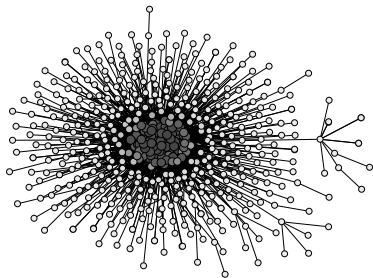


label propagation example



(consensus) partition of European highways with 1174 nodes

label propagation **abstraction**



(left) communities in **Google web graph** with 875 713 nodes

(right) partitioning of **Pennsylvania roads** with 1 087 562 nodes

label propagation conclusions

- (method) **simplest/fastest**/parallel algorithm in literature
- (generality) method for almost **any graph/clustering/use case**
- (literature) reviewed > 150 references & selected **78 references**
- (practice) first network **abstraction** & (**future**) more **applications**

| CONTENTS | |
|------------------------------------|----|
| 1 Label Propagation for Clustering | 9 |
| 1.1 Introduction | 9 |
| 1.2 Related Work | 10 |
| 1.3 Problem Statement | 11 |
| 1.4 Contributions | 12 |
| 1.5 Organization | 13 |
| 2 Preliminaries | 14 |
| 2.1 Graphs | 14 |
| 2.2 Clustering | 15 |
| 2.3 Label Propagation | 16 |
| 2.4 Related Work | 17 |
| 2.5 Organization | 18 |
| 3 Label Propagation for Clustering | 19 |
| 3.1 Introduction | 19 |
| 3.2 Related Work | 20 |
| 3.3 Problem Statement | 21 |
| 3.4 Contributions | 22 |
| 3.5 Organization | 23 |
| 4 Preliminaries | 24 |
| 4.1 Graphs | 24 |
| 4.2 Clustering | 25 |
| 4.3 Label Propagation | 26 |
| 4.4 Related Work | 27 |
| 4.5 Organization | 28 |

| CHAPTER 1 | |
|------------------------------------|----|
| LABEL PROPAGATION FOR CLUSTERING | |
| 1.1 Introduction | 9 |
| 1.2 Related Work | 10 |
| 1.3 Problem Statement | 11 |
| 1.4 Contributions | 12 |
| 1.5 Organization | 13 |
| 2 Preliminaries | 14 |
| 2.1 Graphs | 14 |
| 2.2 Clustering | 15 |
| 2.3 Label Propagation | 16 |
| 2.4 Related Work | 17 |
| 2.5 Organization | 18 |
| 3 Label Propagation for Clustering | 19 |
| 3.1 Introduction | 19 |
| 3.2 Related Work | 20 |
| 3.3 Problem Statement | 21 |
| 3.4 Contributions | 22 |
| 3.5 Organization | 23 |
| 4 Preliminaries | 24 |
| 4.1 Graphs | 24 |
| 4.2 Clustering | 25 |
| 4.3 Label Propagation | 26 |
| 4.4 Related Work | 27 |
| 4.5 Organization | 28 |

| CHAPTER 2 | |
|------------------------------------|----|
| LABEL PROPAGATION FOR CLUSTERING | |
| 2.1 Introduction | 29 |
| 2.2 Related Work | 30 |
| 2.3 Problem Statement | 31 |
| 2.4 Contributions | 32 |
| 2.5 Organization | 33 |
| 3 Preliminaries | 34 |
| 3.1 Graphs | 34 |
| 3.2 Clustering | 35 |
| 3.3 Label Propagation | 36 |
| 3.4 Related Work | 37 |
| 3.5 Organization | 38 |
| 4 Label Propagation for Clustering | 39 |
| 4.1 Introduction | 39 |
| 4.2 Related Work | 40 |
| 4.3 Problem Statement | 41 |
| 4.4 Contributions | 42 |
| 4.5 Organization | 43 |
| 5 Preliminaries | 44 |
| 5.1 Graphs | 44 |
| 5.2 Clustering | 45 |
| 5.3 Label Propagation | 46 |
| 5.4 Related Work | 47 |
| 5.5 Organization | 48 |

| CHAPTER 3 | |
|------------------------------------|----|
| LABEL PROPAGATION FOR CLUSTERING | |
| 3.1 Introduction | 49 |
| 3.2 Related Work | 50 |
| 3.3 Problem Statement | 51 |
| 3.4 Contributions | 52 |
| 3.5 Organization | 53 |
| 4 Preliminaries | 54 |
| 4.1 Graphs | 54 |
| 4.2 Clustering | 55 |
| 4.3 Label Propagation | 56 |
| 4.4 Related Work | 57 |
| 4.5 Organization | 58 |
| 5 Label Propagation for Clustering | 59 |
| 5.1 Introduction | 59 |
| 5.2 Related Work | 60 |
| 5.3 Problem Statement | 61 |
| 5.4 Contributions | 62 |
| 5.5 Organization | 63 |
| 6 Preliminaries | 64 |
| 6.1 Graphs | 64 |
| 6.2 Clustering | 65 |
| 6.3 Label Propagation | 66 |
| 6.4 Related Work | 67 |
| 6.5 Organization | 68 |

thank you!

arXiv:1709.05634v1

Šubelj (2020) Label propagation for clustering, In: *Advances in Network Clustering and Blockmodeling*, pp. 121-150

Lovro Šubelj
University of Ljubljana
lovro.subelj@fri.uni-lj.si
<http://lovro.lpt.fri.uni-lj.si>