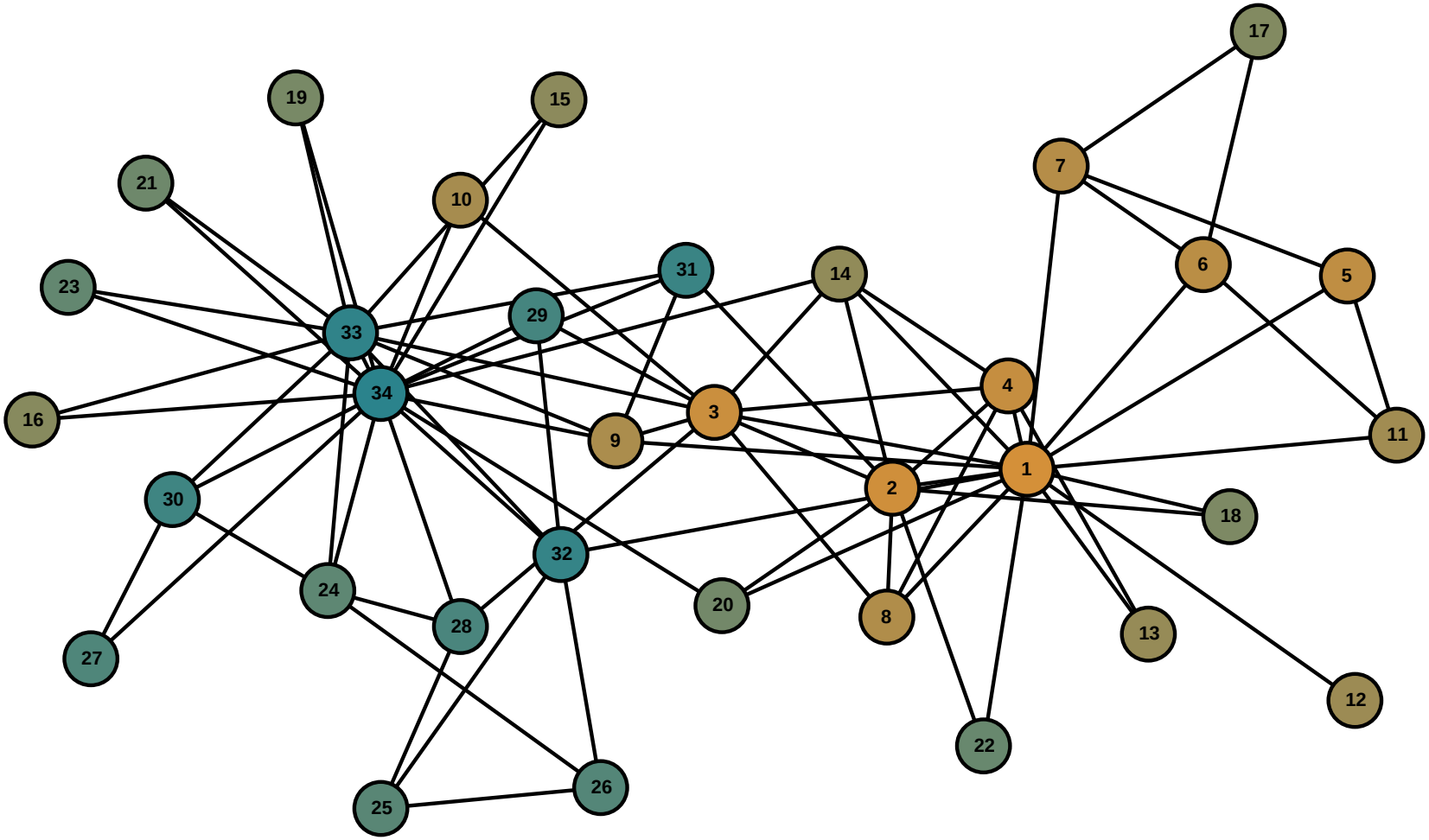
label propagation for clustering

“razvrščanje vozlišč omrežja z izmenjavo oznak”

Lovro Šubelj
University of Ljubljana
Faculty of Computer and
Information Science

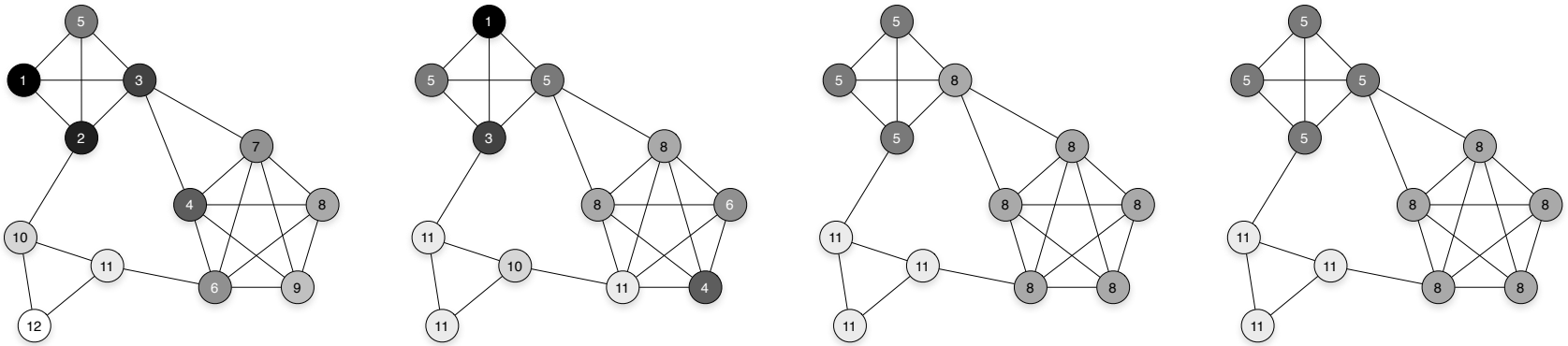
NetSlo ’20

label propagation animation



clusters of nodes represented by **colors**

label propagation method



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

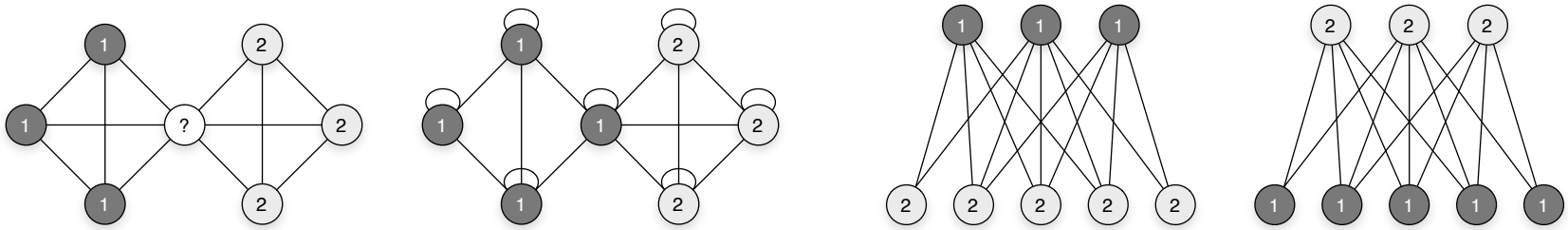
(**initialization**) put each node i in **own cluster** e.g. $c_i = i$

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

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

(**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

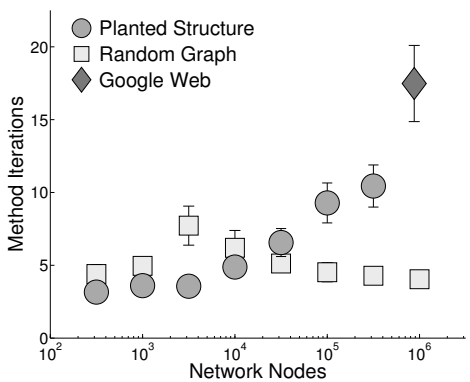
(**optimization**) label propagation is equivalent to **Potts model**

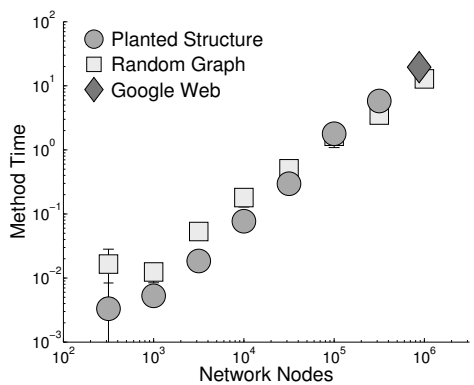
$$\mathcal{F}(\{c\}) = \sum_{ij} A_{ij}\delta(c_i, c_j)$$

(**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.





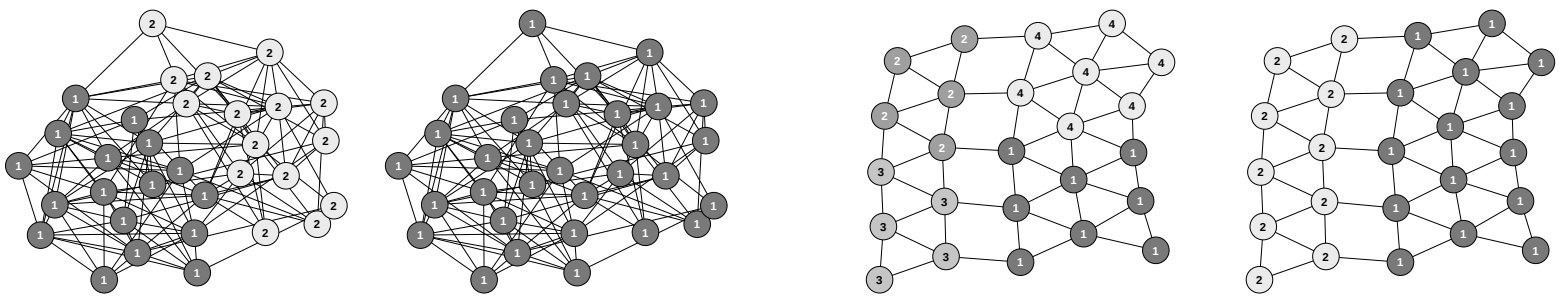
label propagation advances

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

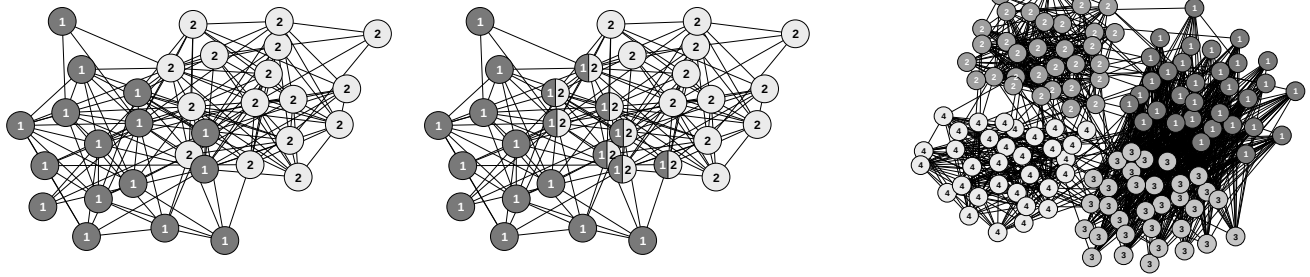
$$c_i = \underset{c}{\operatorname{argmax}} \sum_j \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_j, c)$$

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

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



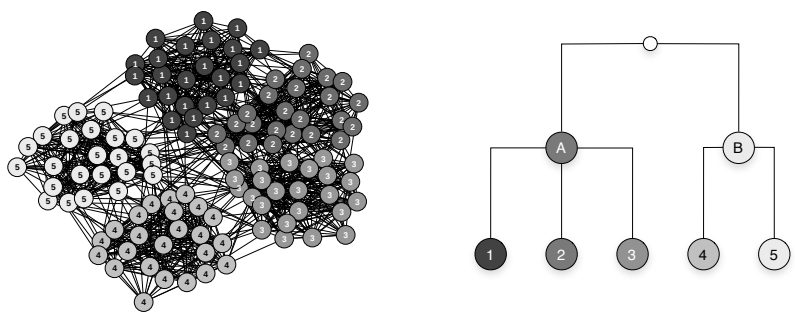
label propagation clusters



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

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

(bottom) hierarchical detection of nested clusters' dendrogram



label propagation networks

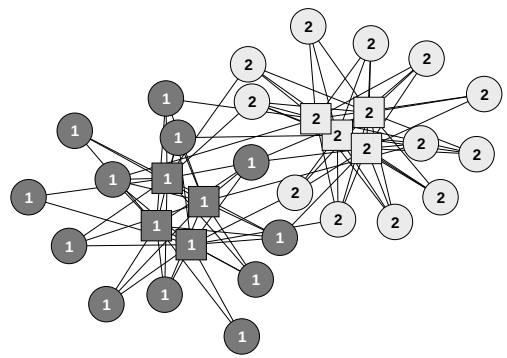
(**weights**) **straightforward** for weighted multigraphs ↑

(**directions**) there is **no general** method for directed graphs

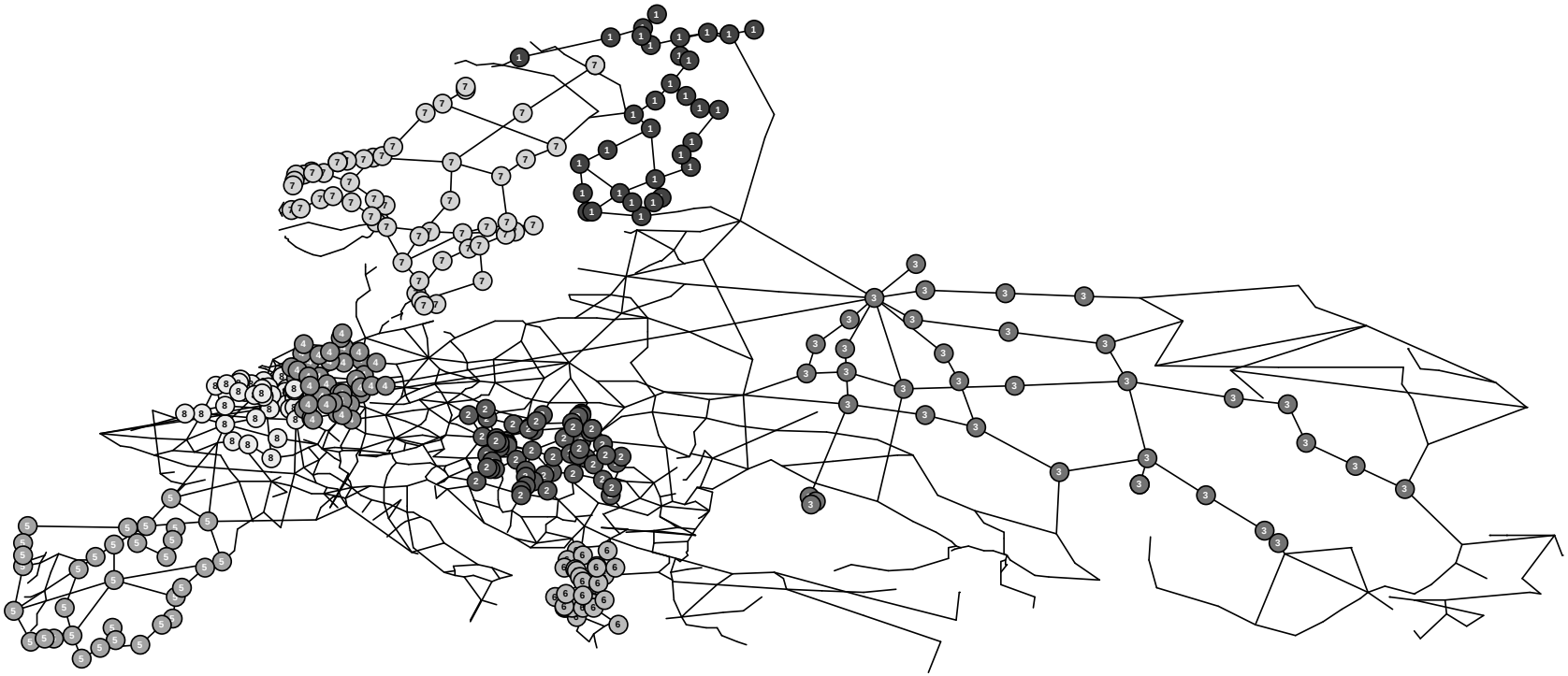
(**signs**) signed graphs with m_p **positive** & m_n **negative** links

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

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

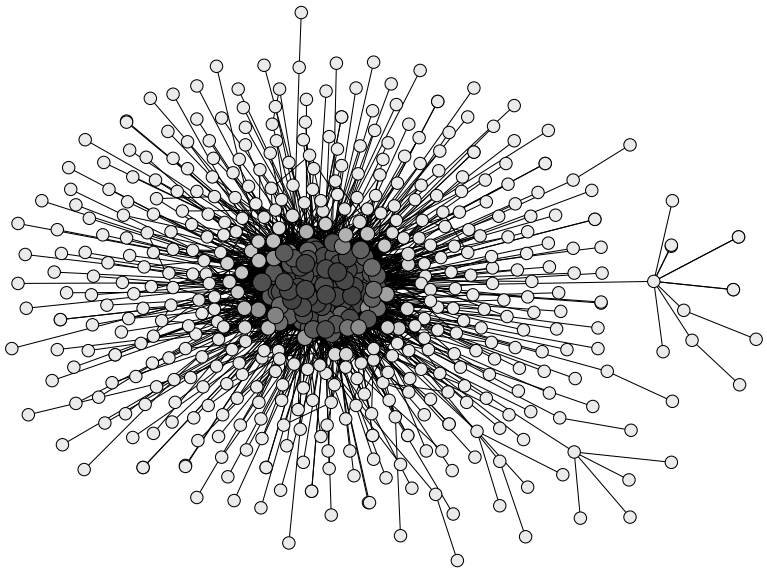


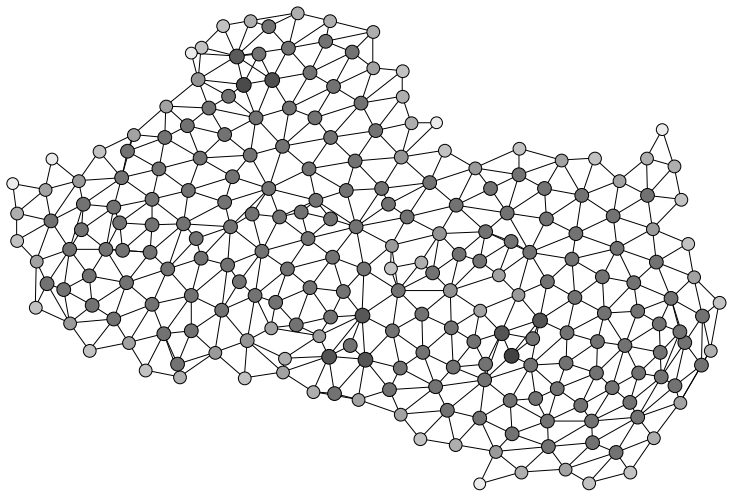
label propagation example



(**consensus**) partition of **European highways** with 1 174 nodes

label propagation abstraction





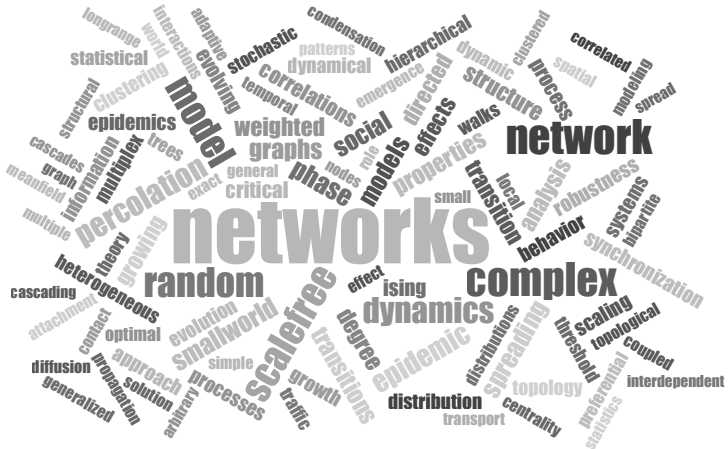
(**left**) communities in **Google web graph** with 875 713 nodes

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

label propagation applications

(**recommendation**) **people you may know** on Facebook*

(**compression**) **degrees of separation** between Facebook users†



(**bibliometrics**) clusters of **papers published** in Physical Review E

---

* Ugander & Backstrom (2013) Balanced label propagation, In: *Proceedings of WSDM '13*, pp. 507-516

† Boldi et al. (2011) Layered label propagation, In: *Proceedings of WWW '11*, pp. 587-596

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**

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