

**Antonio Maria Fiscarelli**

Matthias R. Brust

Grégoire Danoy

Pascal Bouvry

# A Memory-Based Label Propagation Algorithm for Community Detection

COMPLEX NETWORKS 2018

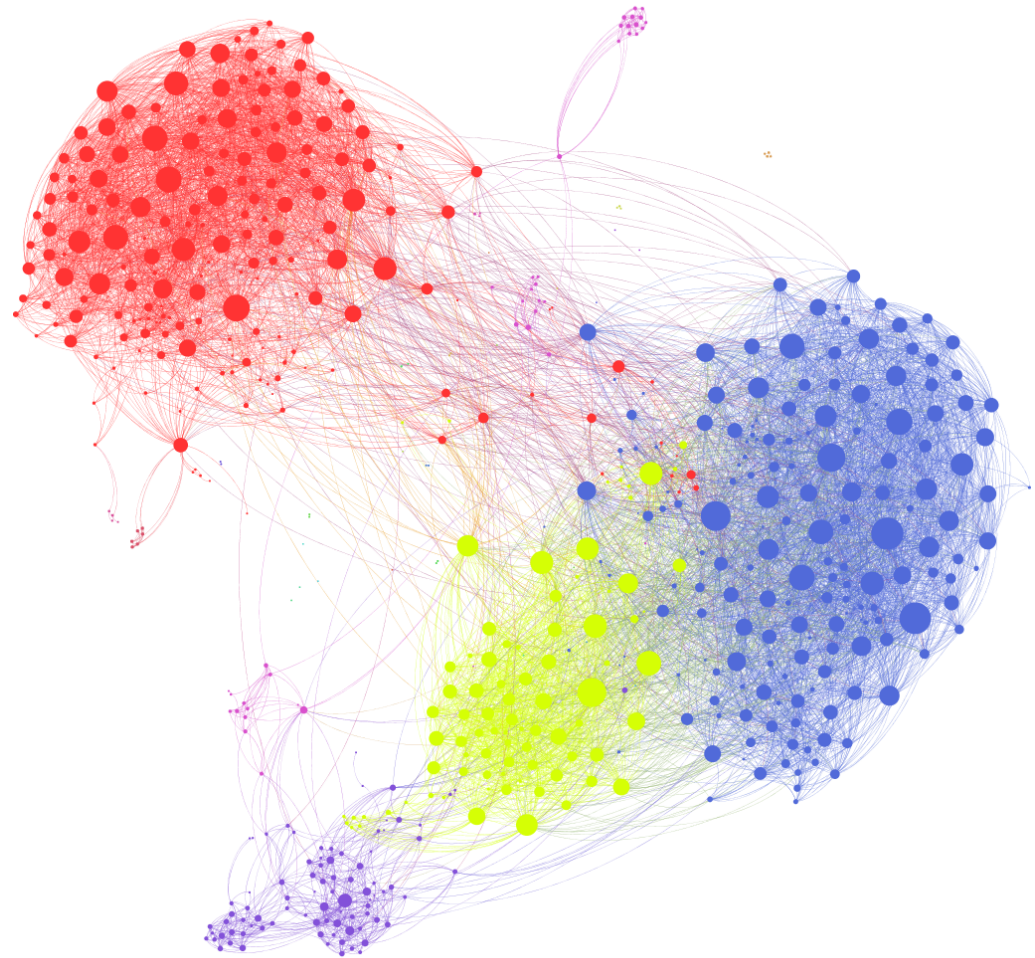


- Community Detection
- Literature review
- MemLPA
- Results
- Conclusions
- Future work



# Community Detection

- Given a graph  $G=\{V, E\}$ , community detection consists of grouping related vertices that show:
  - High intra-cluster connectivity
  - Low inter-cluster connectivity



- Community detection is relevant in many fields
  - **Social Sciences**
    - Friendship networks
    - Citation/Collaboration networks
    - Epidemic spreading (Fake news, Viral topics)
  - **Multi-agent systems**
    - MANET
    - Swarms of UAVs

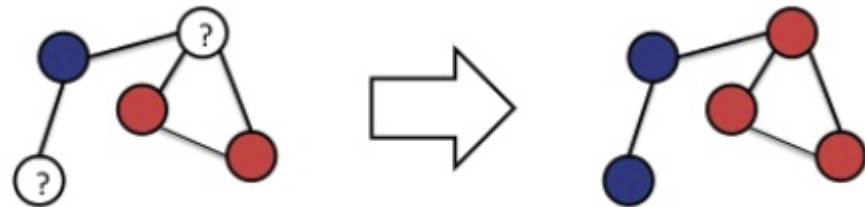
- **Modularity optimization methods**
  - Betweenness [1]
  - Fast greedy [2]
  - Louvain [3]
- **Random walks-based methods:** uses random walks to discover community structure
  - Walktrap [4]
  - Infomap [5]
- **Spectral methods**
  - Leading Eigenvector [6]

## Label Propagation

A unique label is assigned to every node and they use a majority rule to agree on a common label [7]

- **Advantages:**

- Decentralized
- Near-linear complexity
- Parameter free



- **Disadvantages:**

- Outperformed by more sophisticated algorithms
- Gets stuck in local optima
- Can oscillate between two different configurations

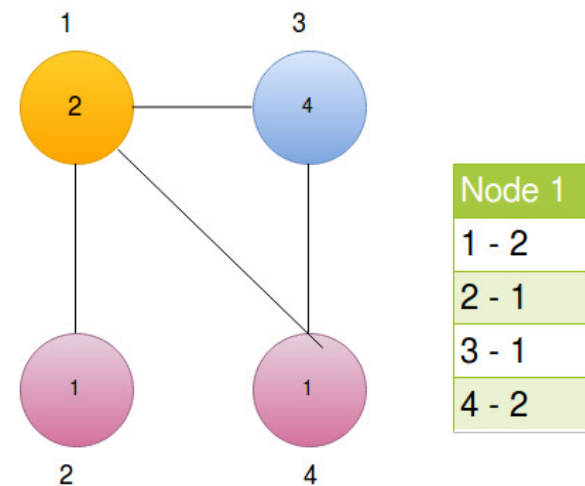
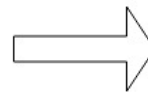
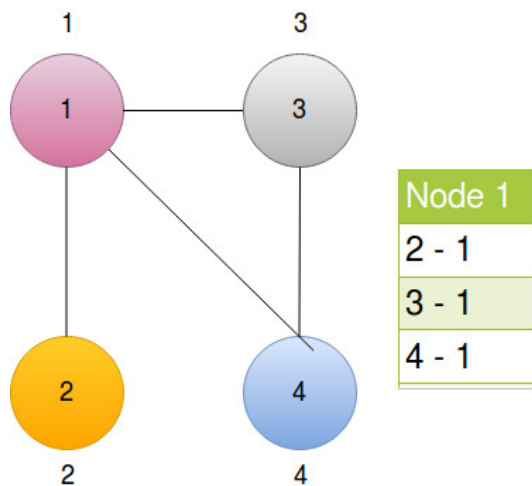
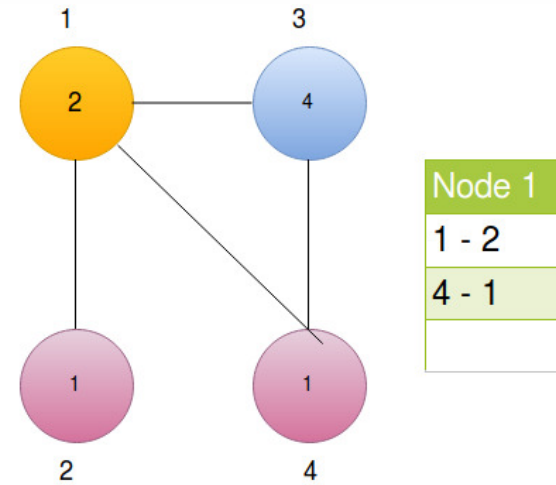
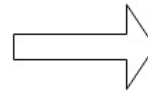
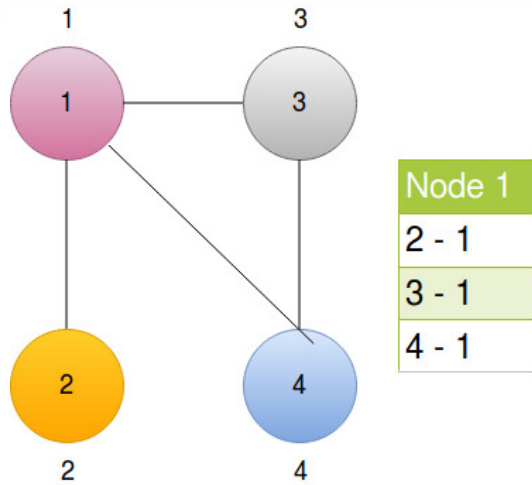
## Different LPA variations [8-13]

- **Initialization**
  - Unique labels
  - Limited labels
  - Unlabeled nodes
  -
- **Update**
  - Synchronous
  - Asynchronous
  - Conditional
- **Decision rule**
  - Neighborhood strength
  - Propagation attenuation
- **Termination criterion**
  - Convergence
  - Modularity improvement
  - Active nodes
  - Scarcity of updates

## Different LPA variations [8-13]

- **Initialization**
  - *Unique labels*
  - Limited labels
  - Unlabeled nodes
- **Update**
  - *Synchronous*
  - Asynchronous
  - Conditional
- **Decision rule**
  - *Neighborhood strength*
  - Propagation attenuation
- **Termination criterion**
  - Convergence
  - Modularity improvement
  - *Active nodes*
  - Scarcity of updates





---

**Algorithm 1:** MemLPA

---

**Input** : Graph  $G(N, E)$

**Output:** Communities  $C$

```
1  $AL \leftarrow N$  //Initialize active list
2 for  $n \in N$  do
3    $c_n \leftarrow l_n$  //Assign unique label to nodes
4    $L_n \leftarrow \emptyset$  //Initialize label lists
5 end
6 while  $AL \neq \emptyset$  do
7   for  $n \in AL$  do
8      $C_n \leftarrow CollectLabels(Neigh(n))$  ;
9      $L_n \leftarrow UpdateLabelList(C_n)$  ;
10     $L_n \leftarrow \{l_n^m \in L_n, m \in N \mid |mean(L_n) - sd(L_n)| \leq l_n^m\}$ 
11     $c_n \leftarrow ApplyRule(L_n)$  ;
12  end
13   $AL \leftarrow UpdateActiveList(AL)$ 
14 end
```

---

## Simulations

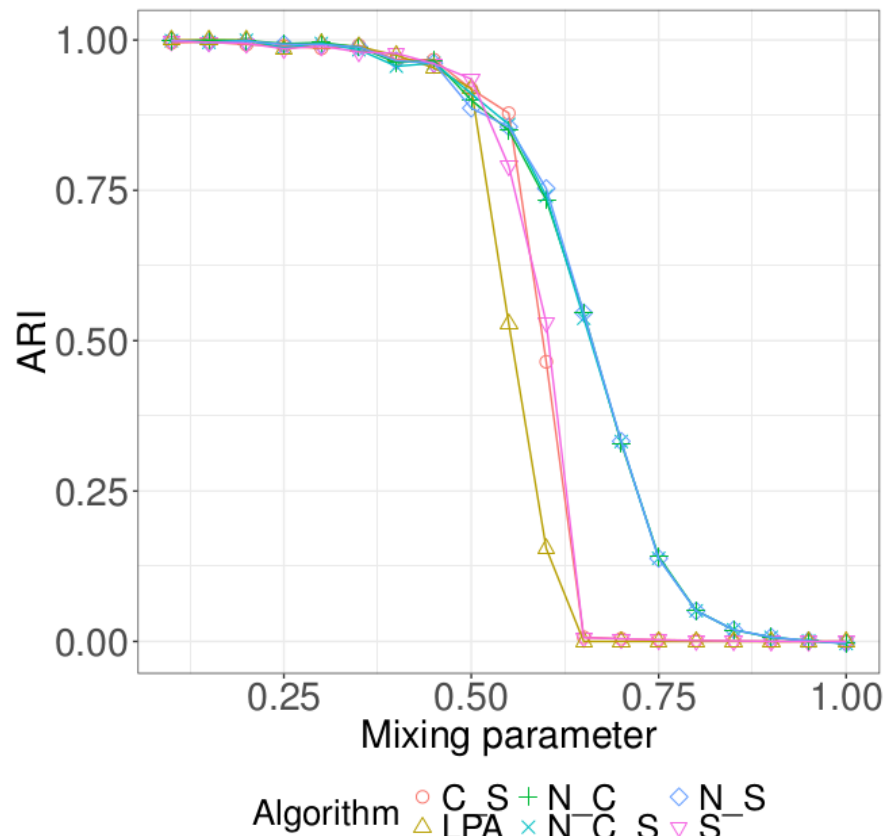
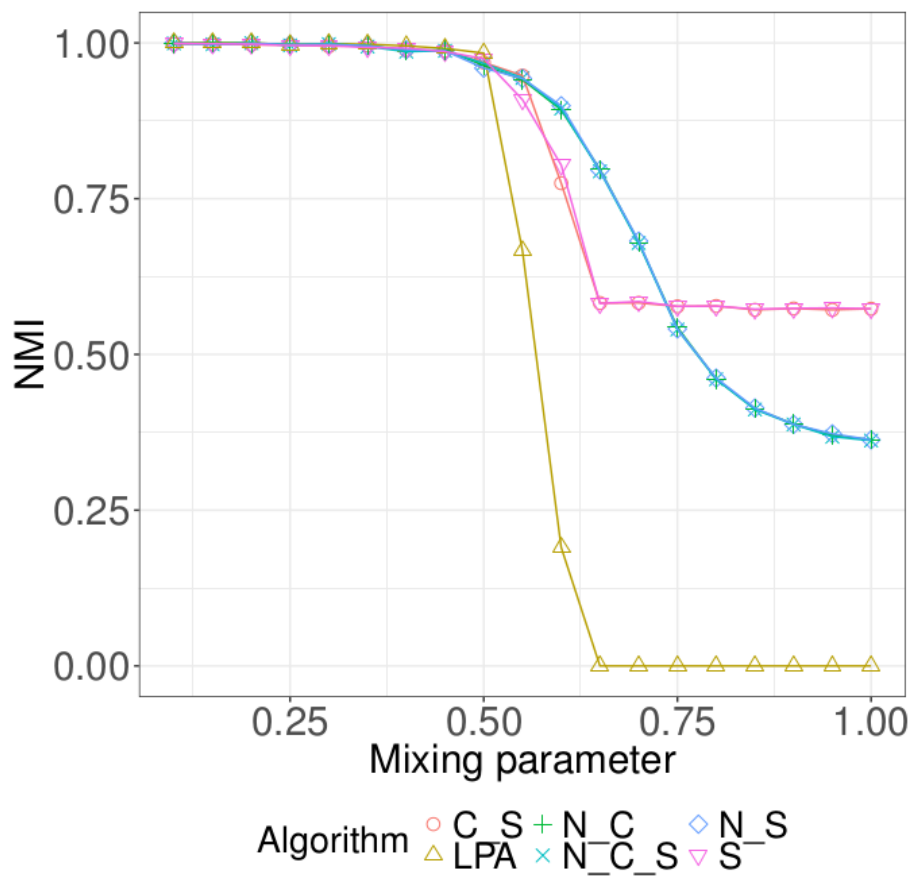
- Artificial networks:
  - Lancichinetti-Fortunato-Radicchi

#nodes	average degree	directed	weighted	#communities	subcommunities
1000	10	no	no	[10-50]	no

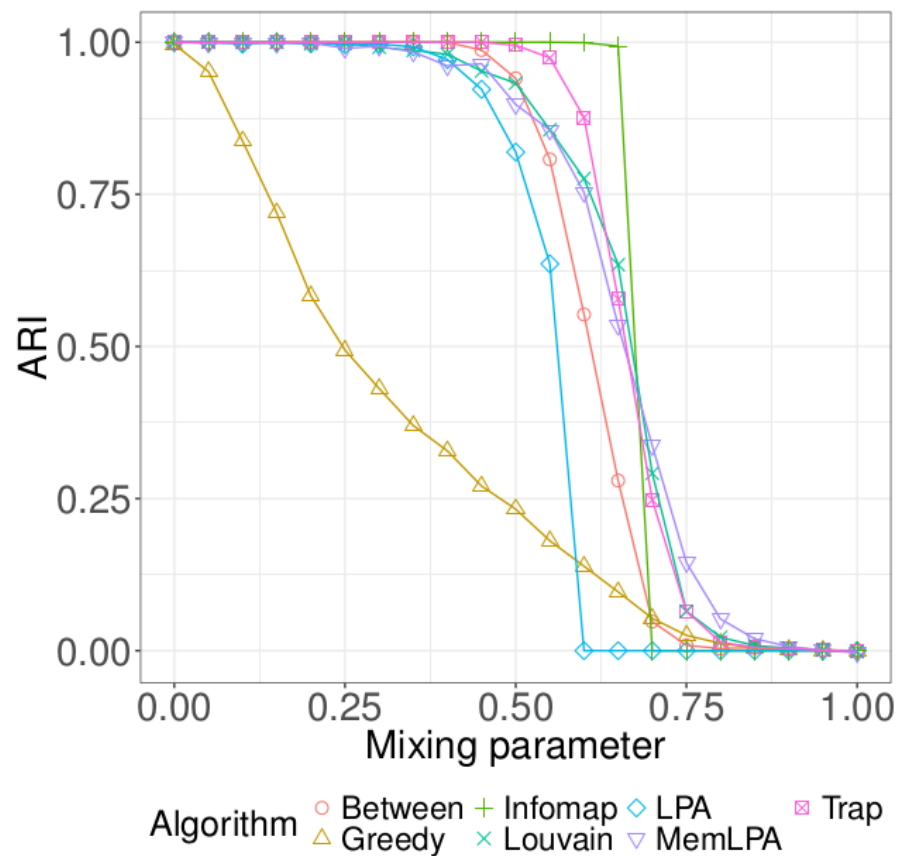
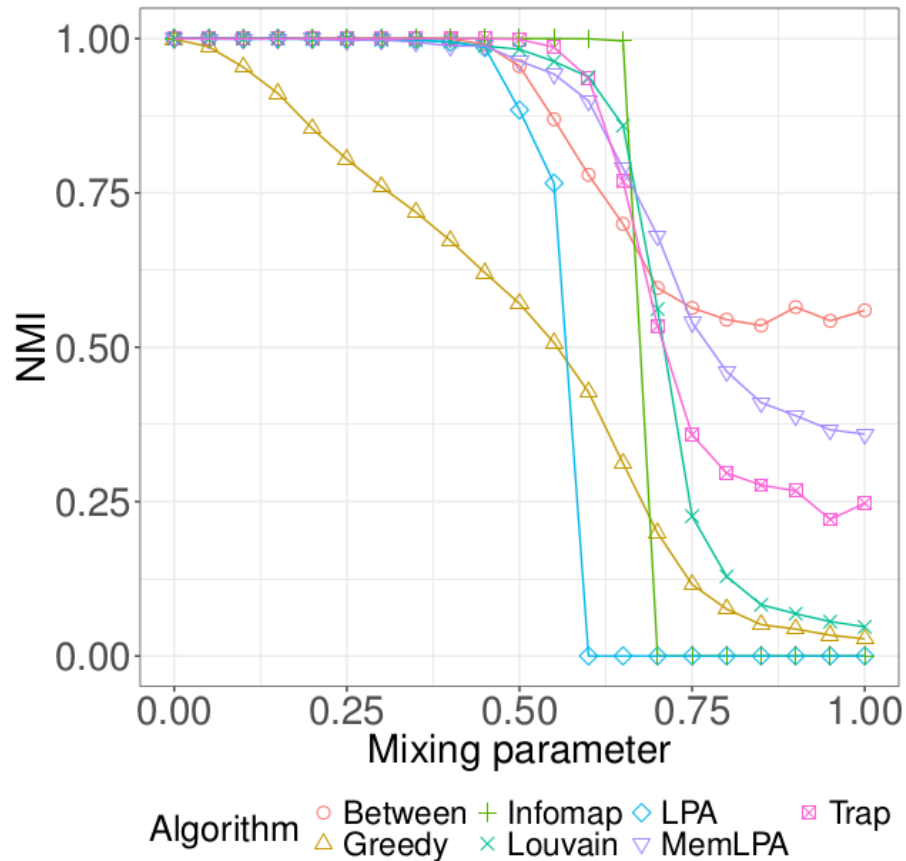
- Real-world networks:

	#nodes	#edges	directed	weighted
karate	34	78	no	yes
UKfaculty	81	817	yes	yes
mail	184	2116	yes	no
dolphins	62	159	yes	no
jazz	198	2742	yes	no
USAirports	755	23473	yes	yes

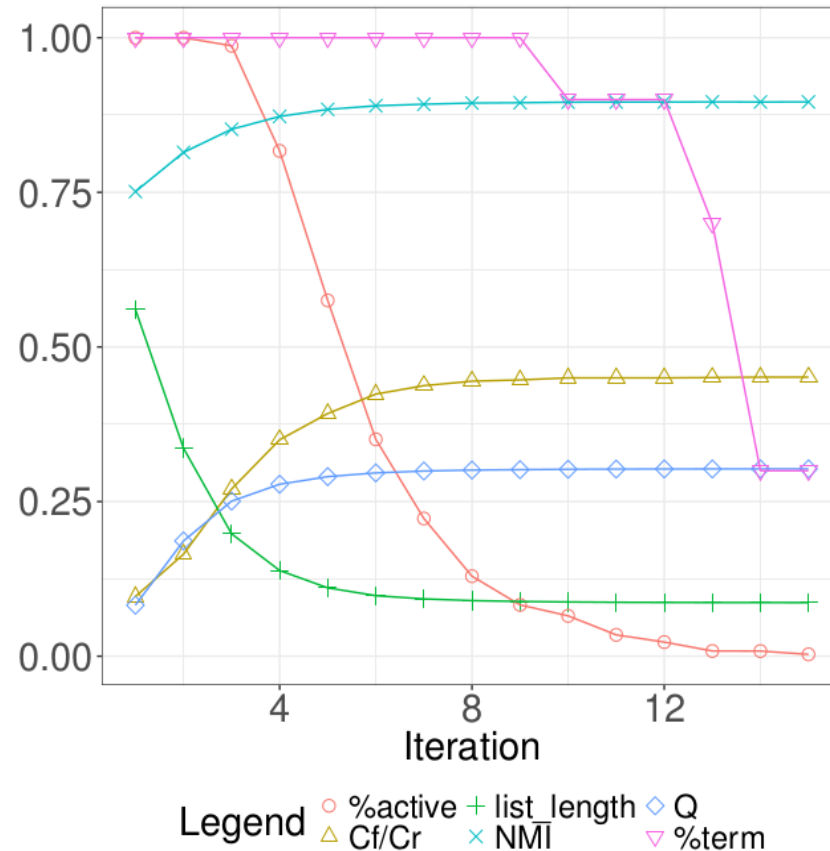
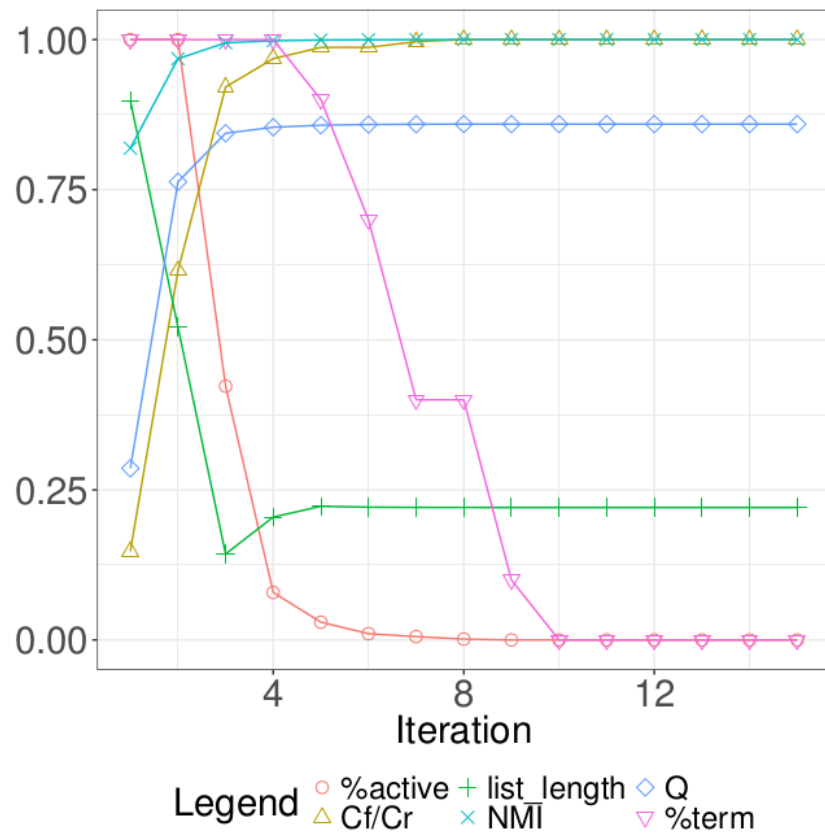
# Results



# Results

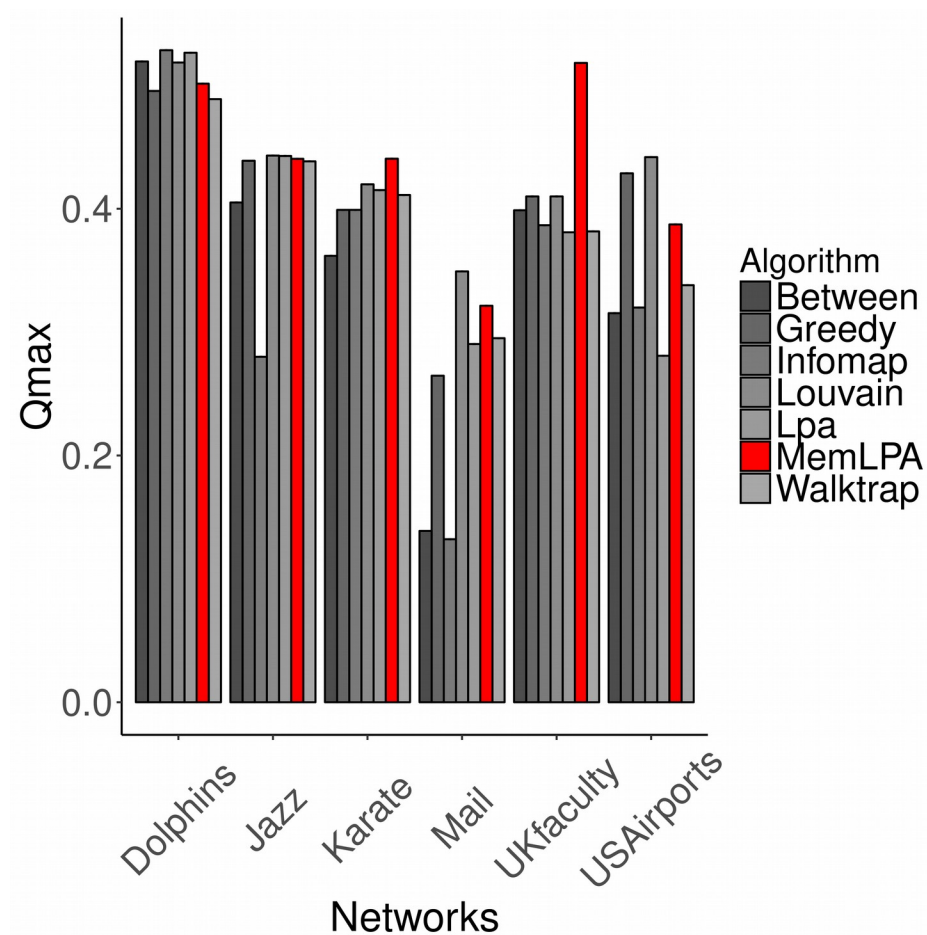
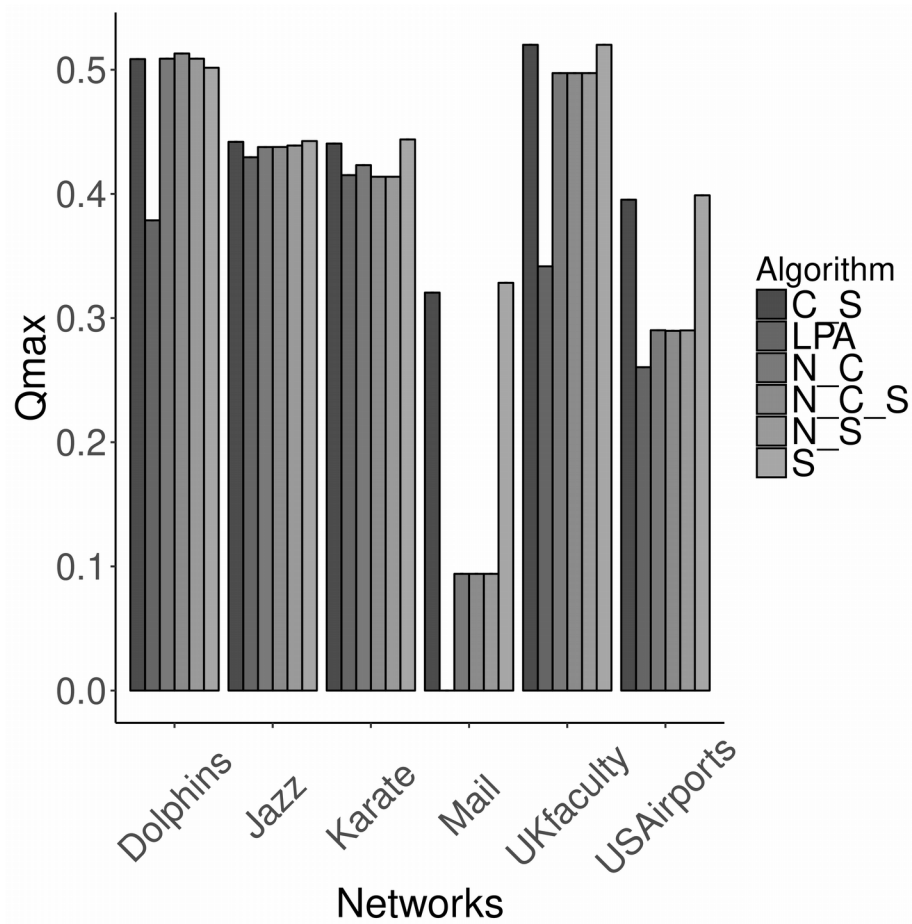


# Results





# Results



- The use of memory increases performance
- The use of memory prevents a single label to flood the whole network

- Study correlation with network structure
- Study topological properties of community found
- New real-world networks

- [1 ] Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. *Physical review* E69(2) (2004)
- [2] Clauset, A., Newman, M.E., Moore, C.: Finding community structure in very large networks. *Physical review* E70(6) (2004)
- [3] Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008(10) (2008)
- [4] Pons, P., Latapy, M.: Computing communities in large networks using random walks. In: *ISCIS*, vol. 3733, pp. 284–293 (2005)
- [5] Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 105(4), 1118–1123 (2008)
- [6] Reichardt, J., Bornholdt, S.: Statistical mechanics of community detection. *Physical Review* E74(1) (2006)
- [7] Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. *Physical review* E76(3) (2007)
- [8] Barber, M.J., Clark, J.W.: Detecting network communities by propagating labels under constraints. *Physical Review* E80(2) (2009)
- [9] Leung, I.X., Hui, P., Lio, P., Crowcroft, J.: Towards real-time community detection in large networks. *Physical Review* E79(2009)
- [10] Liu, X., Murata, T.: Advanced modularity-specialized label propagation algorithm for detecting communities in networks. *Physica A: Statistical Mechanics* 389(7), 1493–1500 (2010)
- [11] Subelj, L., Bajec, M.: Unfolding communities in large complex networks: Combining defensive and offensive label propagation for core extraction. *Physical Review* E83(3) (2011)
- [12] Xie, J., Szymanski, B.K.: Community detection using a neighborhood strength driven label propagation algorithm. *arXiv preprint arXiv:1105.3264* (2011)
- [13] Dongen, S.: A cluster algorithm for graphs (2000)

# Questions

