# Generating Scaled Replicas of Real-World Complex Networks

CHRISTIAN L. STAUDT, MICHAEL HAMANN, ILYA SAFRO,
ALEXANDER GUTFRAIND, HENNING MEYERHENKE
The 5th International Workshop on Complex Networks & their Applications
Nov 30 - Dec 02 2016 Milano

Our Contribution: We…

- introduce the **ReCoN** (for **Re**plication of **Co**mplex **N**etworks) generator
- experimentally evaluate (in comparison with competing generative models)
  - generating **realistic randomized replicas**
  - generating **realistically scaled-up versions**
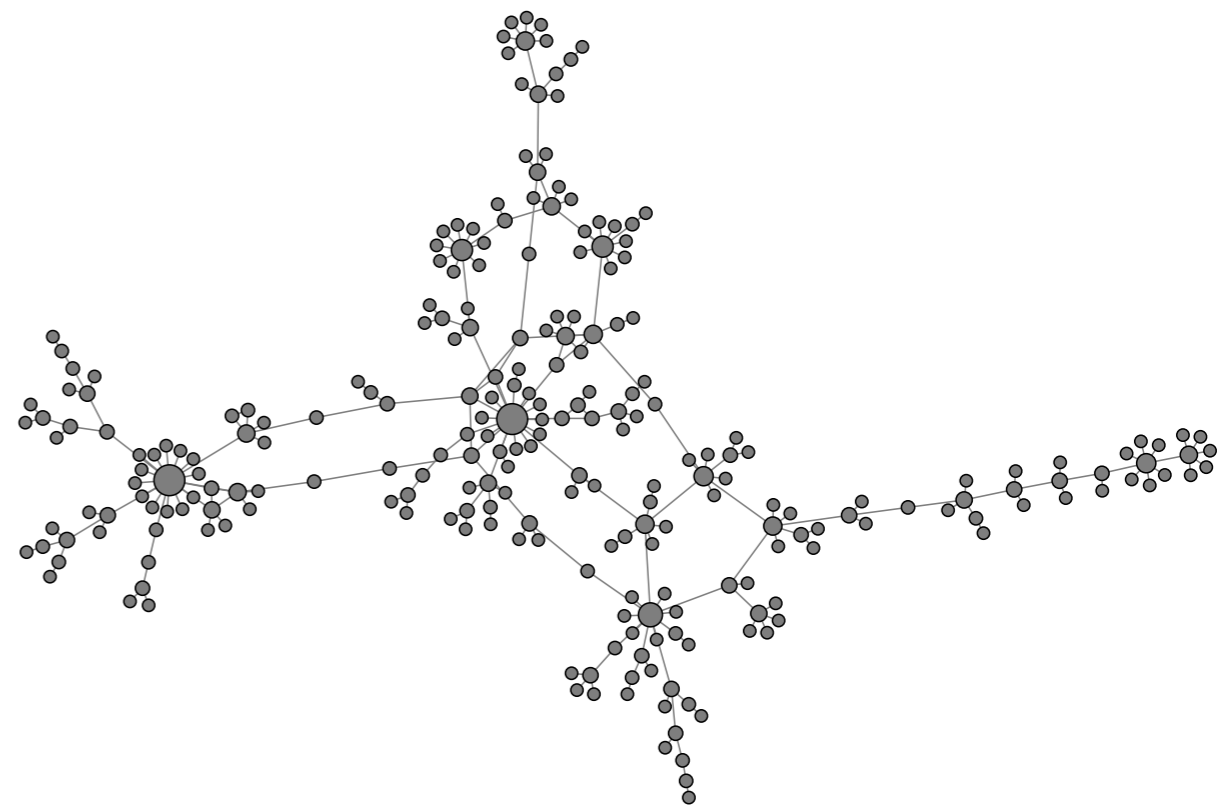
Motivation

Engineering network algorithms? Consider this argument…
[J. Kunegis, http://konect.uni-koblenz.de/]

- task: determine wether algorithm X performs better than algorithm Y
- result: X performs better on 6 out of 10 datasets
- null hypothesis: X and Y win with equal probability on any data set
- under the null hypothesis, probability of obtaining the result above is **17 %**
- **65 datasets needed** for statistically significant result (assuming a p-value of ≤ 0.05) for a 60 % outcome

Why is realism important?

- algorithm performance may **strongly depend on structural properties** of the network

Why synthetic networks?

- enough relevant real network data may be **hard to obtain**, **proprietary**, **sensitive**, or at **wrong scale**
- synthetic graphs often necessary to test scalability and effectiveness of algorithms

Scenarios for ReCoN
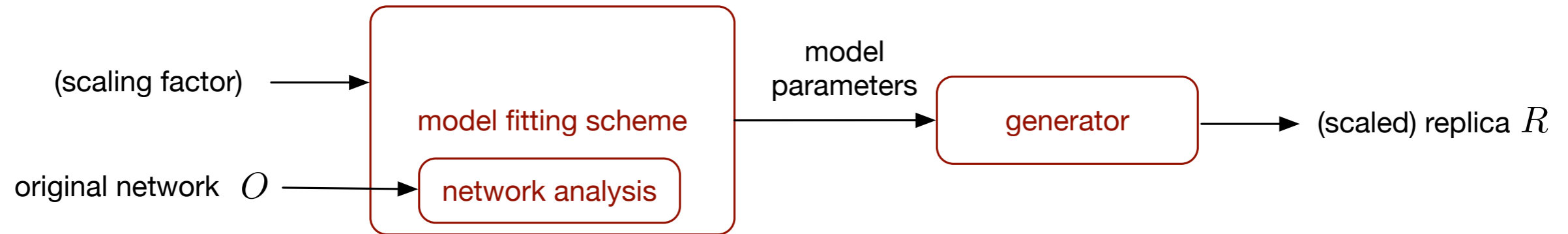
A: Obfuscation
- given a real network that cannot be freely shared, generate a randomized/ obfuscated replica with (statistically) similar properties
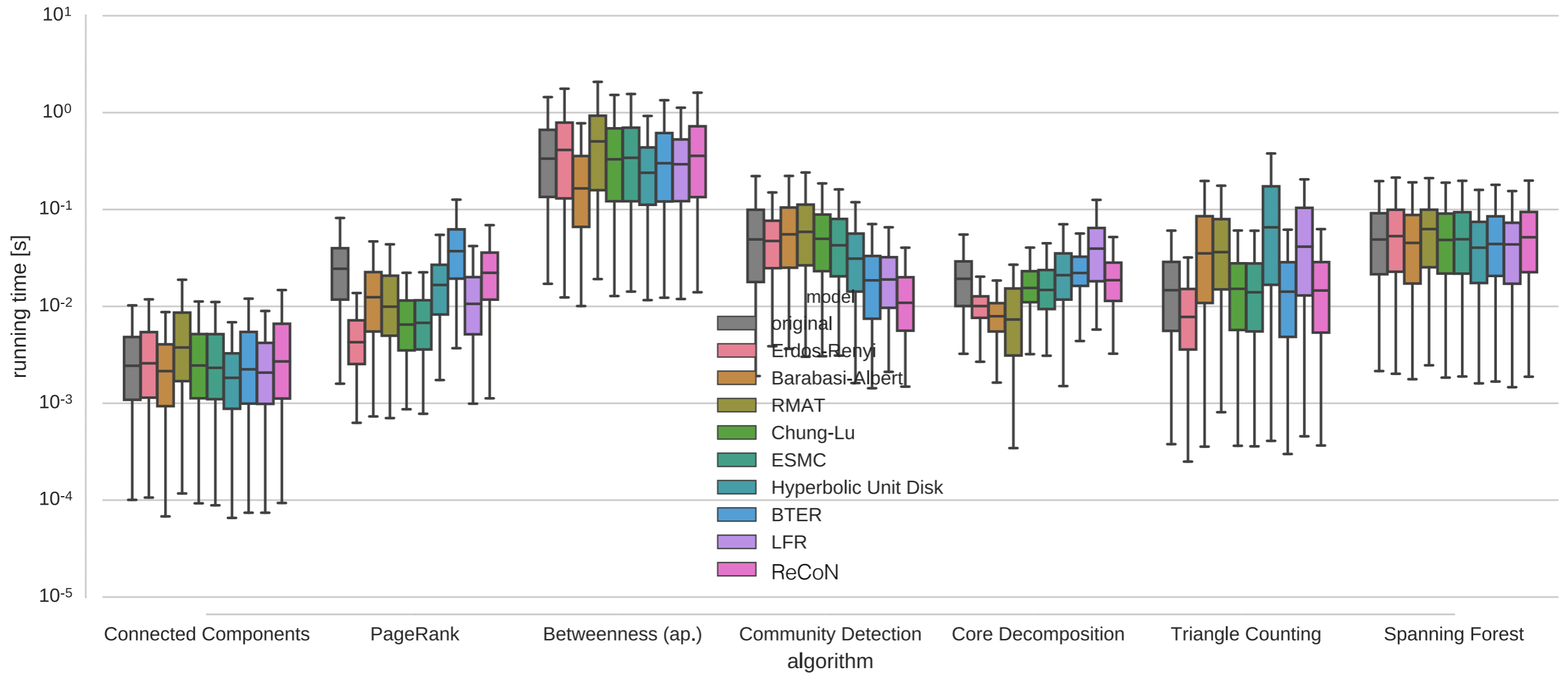
B: Scaling
- given a real network with n nodes, generate a scaled-up network with x n nodes with realistically scaled properties
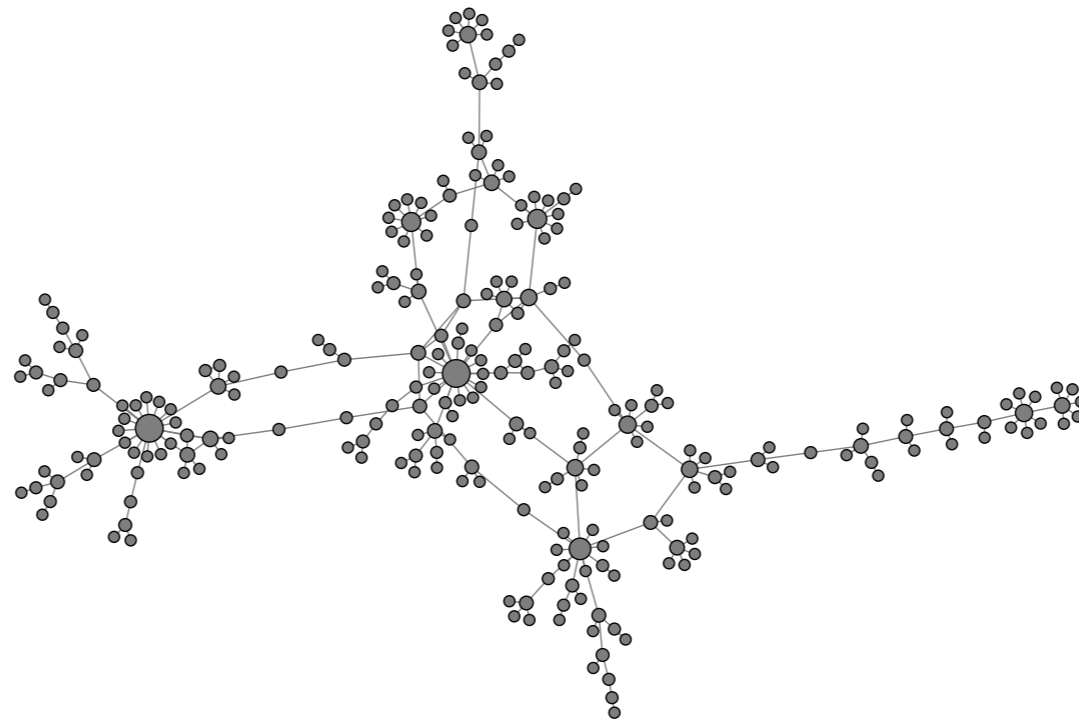
Performance Goal
- generate millions of edges quickly in practice
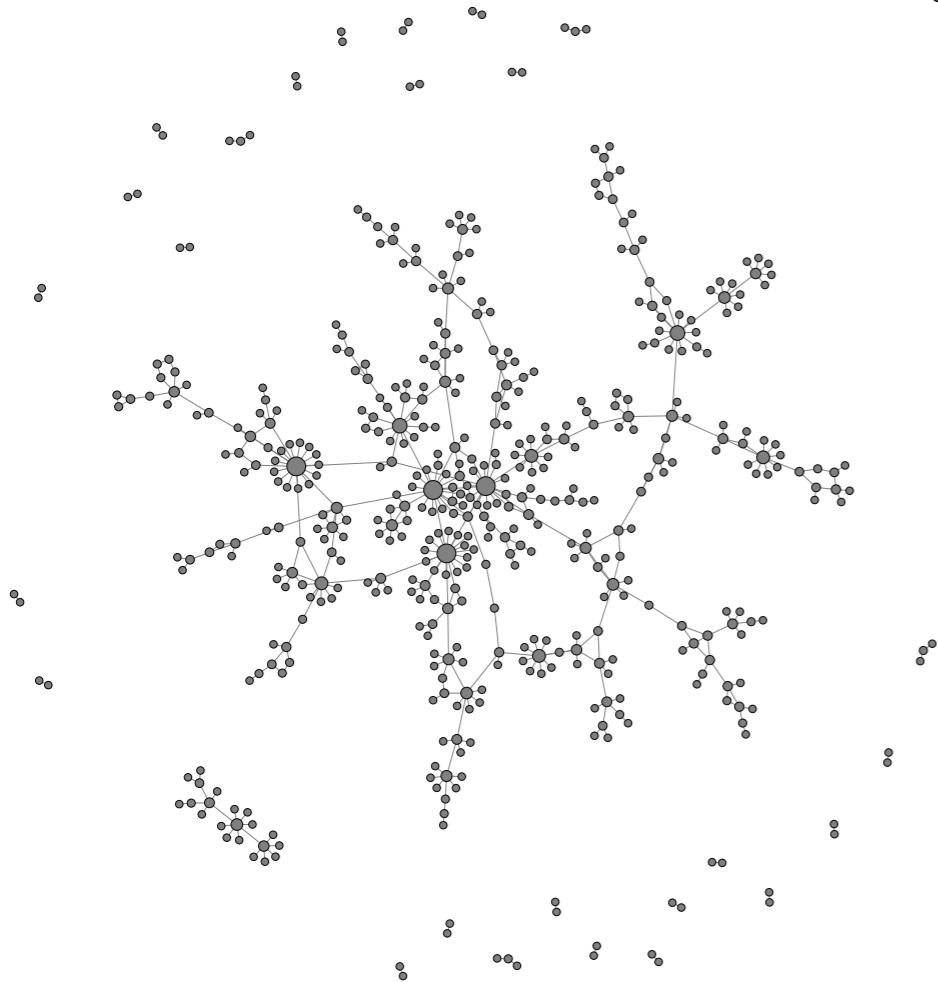
Running Time Replication

Are algorithm running times obtained on synthetic graphs representative for those on real-world inputs?
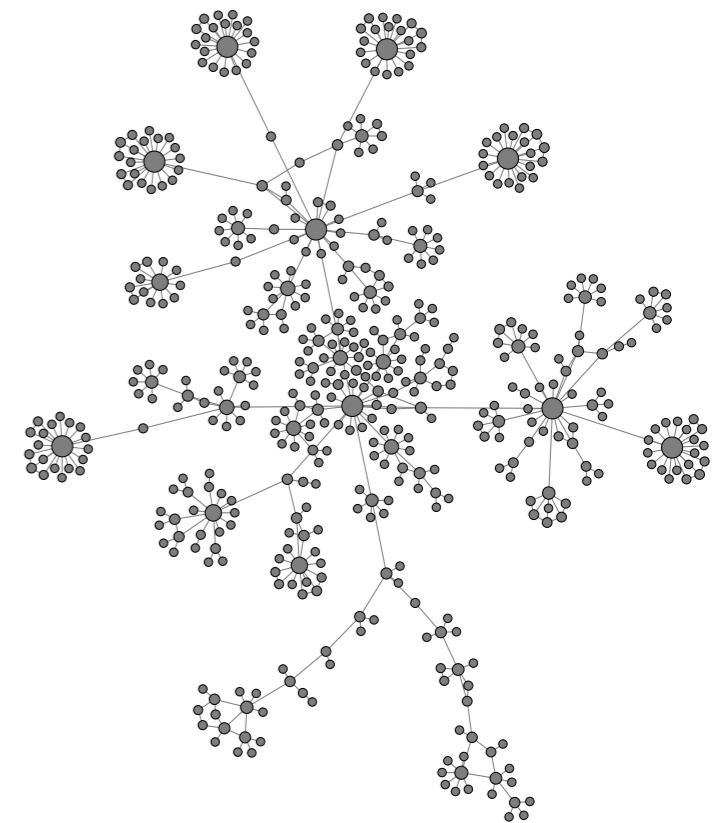
Example: Scaled Replica

epidemiological contact network used in HIV research [Potterat et al. '02]
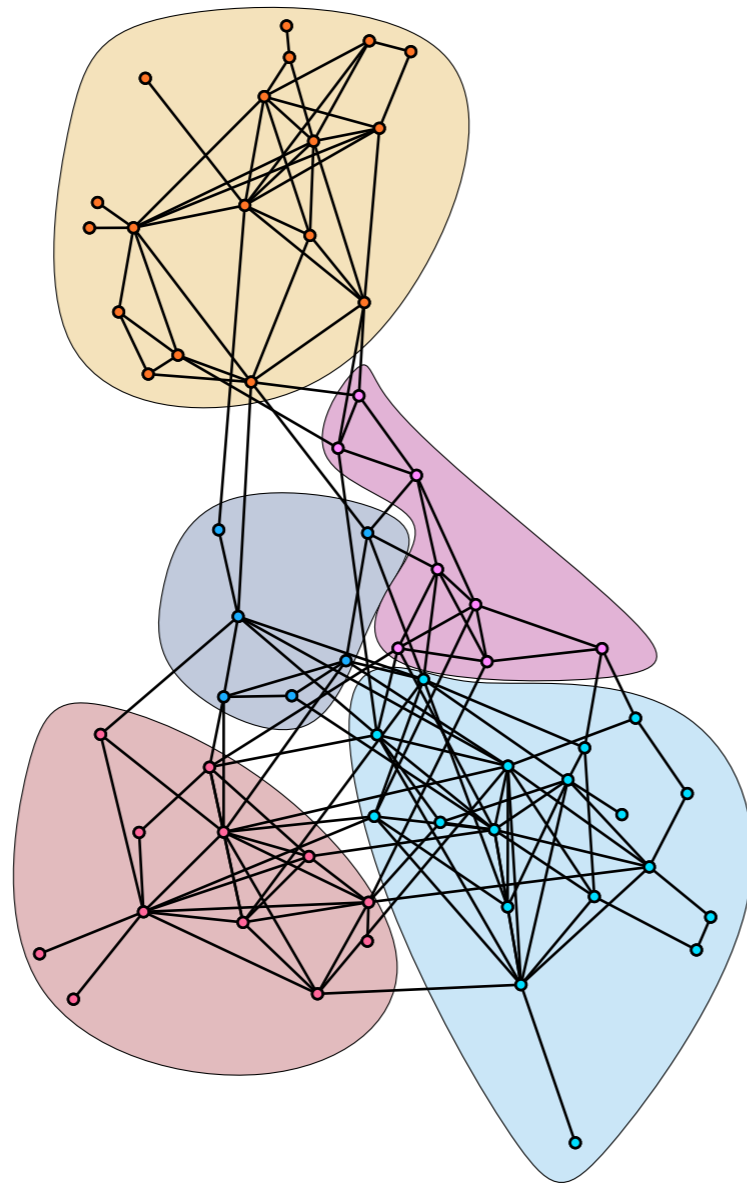
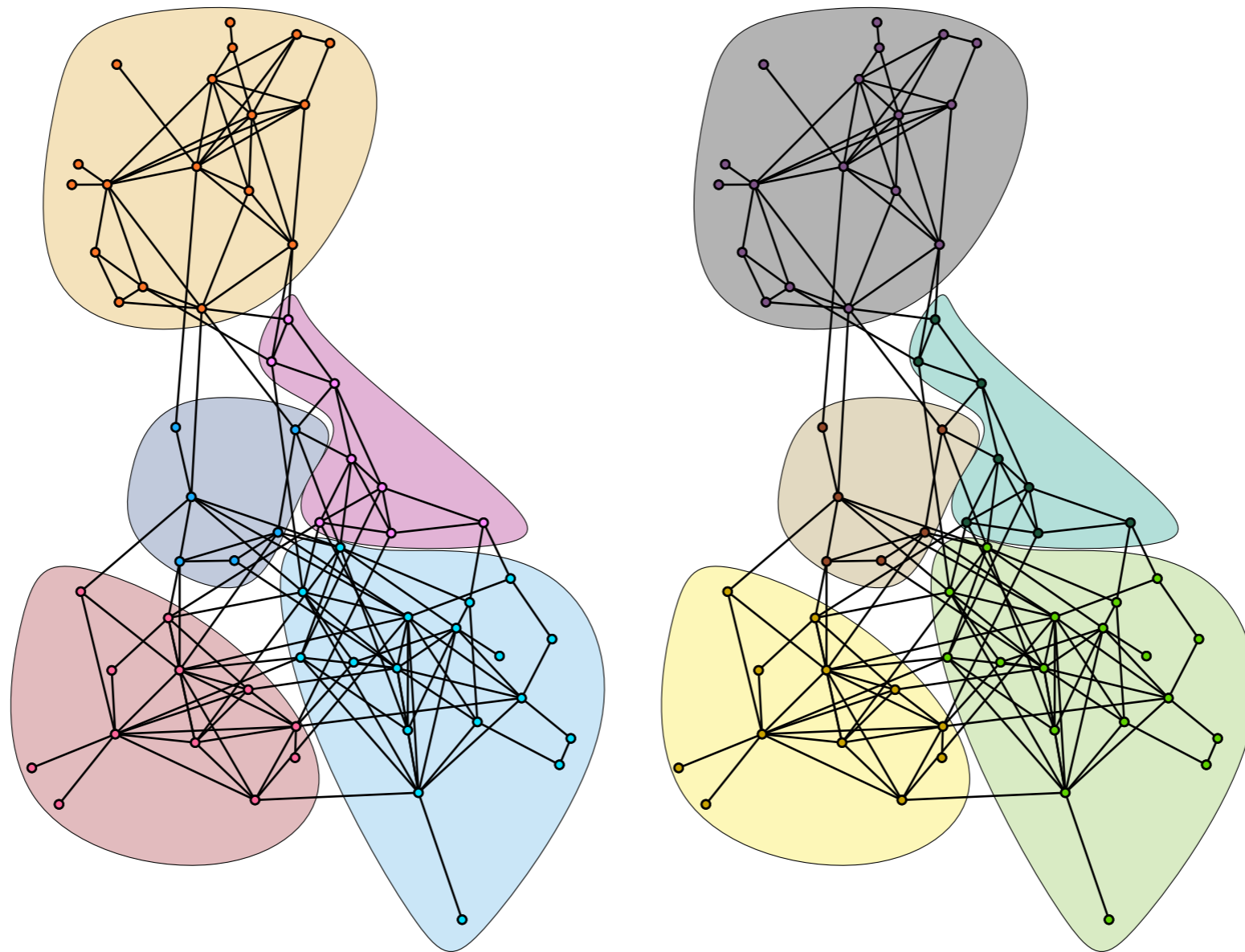scale-2 replica produced by ReCoN generator

sample from scale-200k replica produced by ReCoN generator
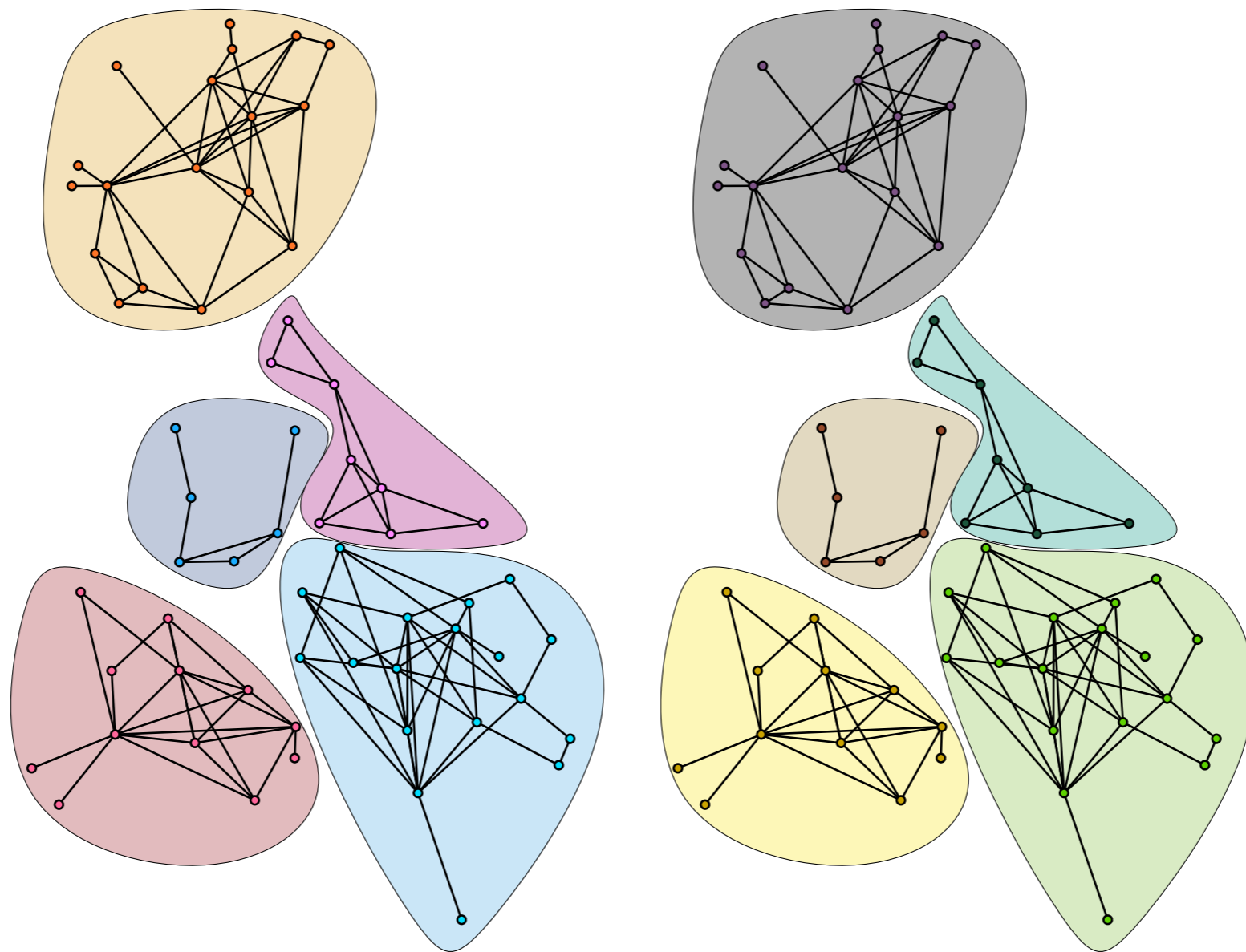
ReCoN Algorithm Example

original network and community structure

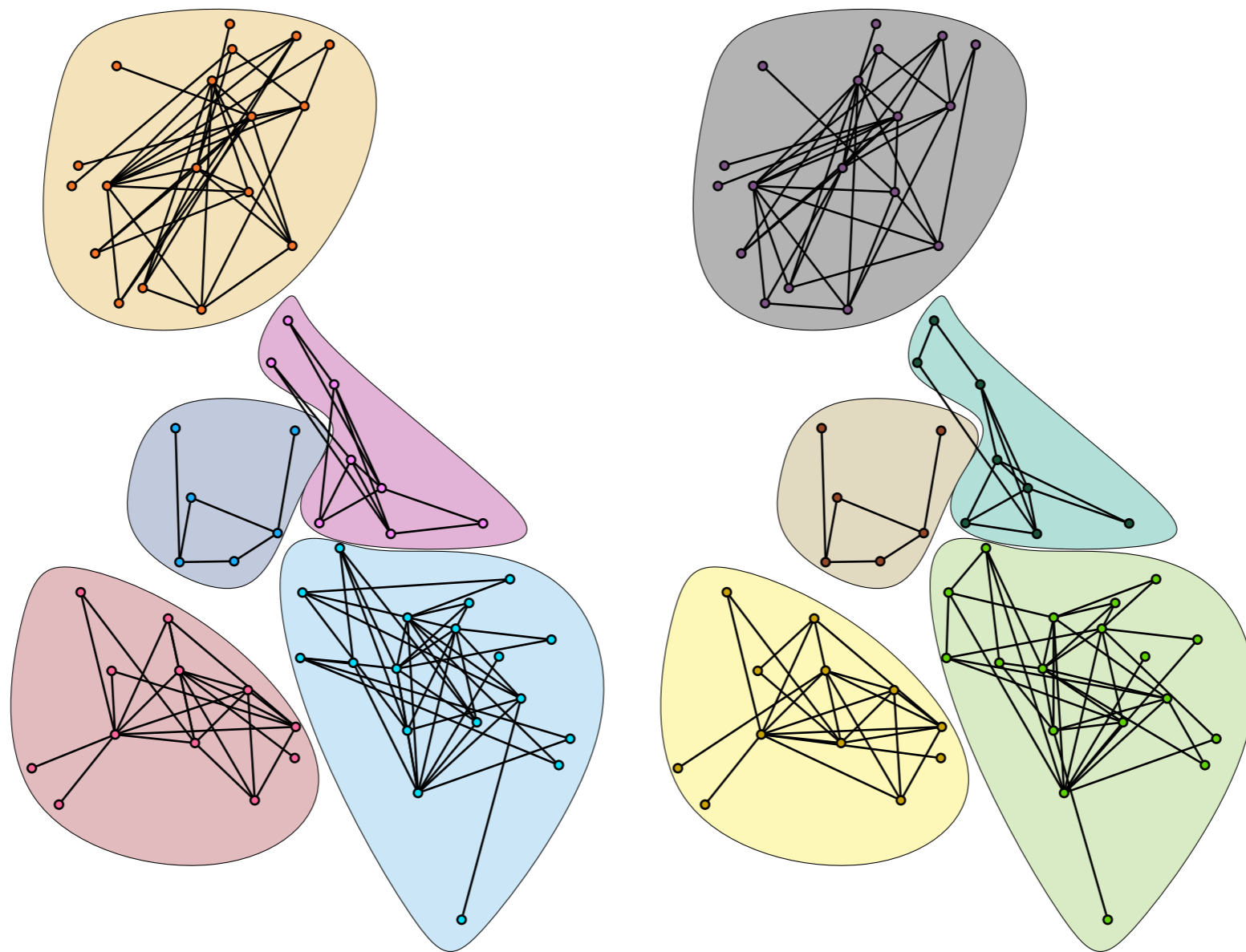# ReCoN Algorithm Example

2 copies

ReCoN Algorithm Example
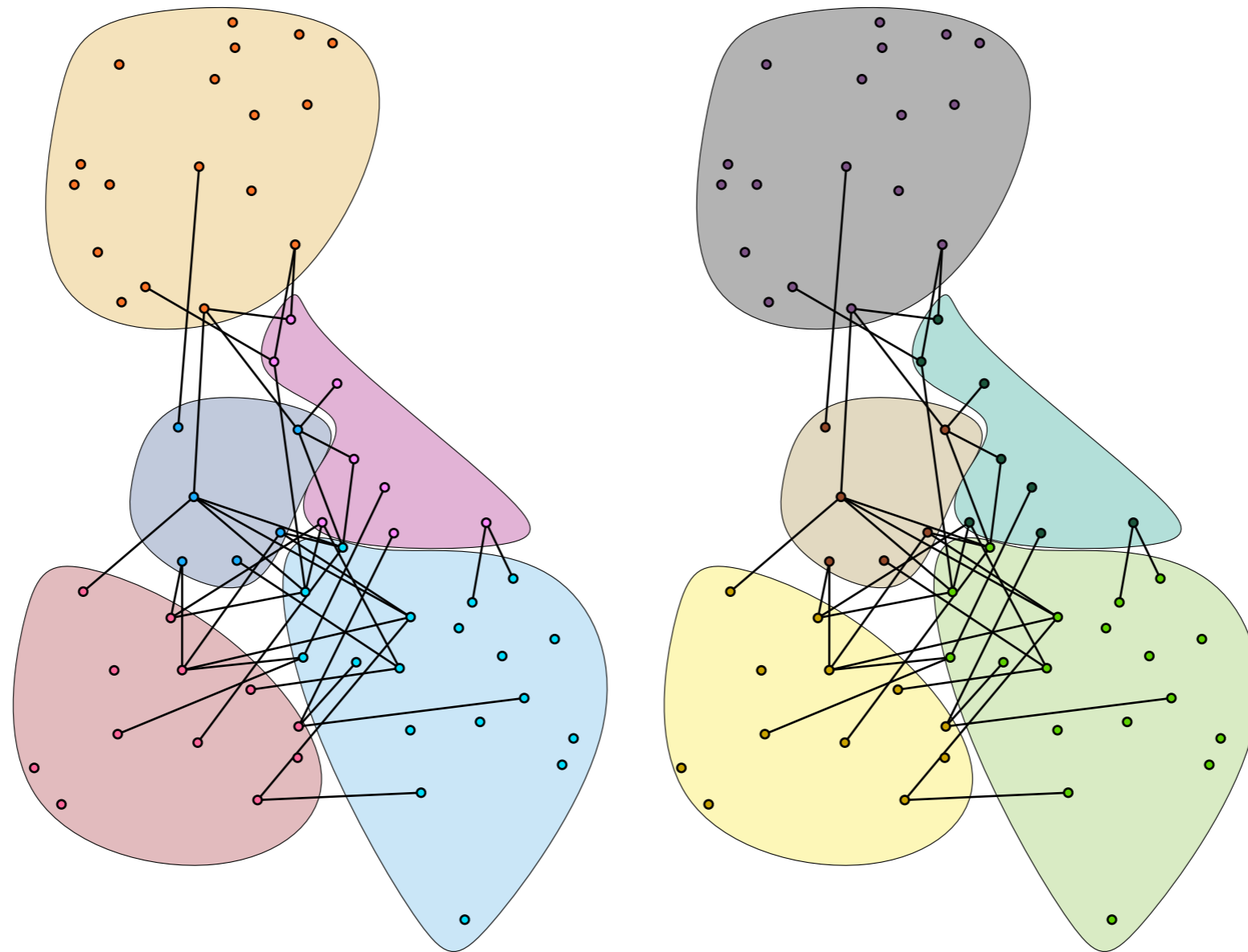

intra-community edges
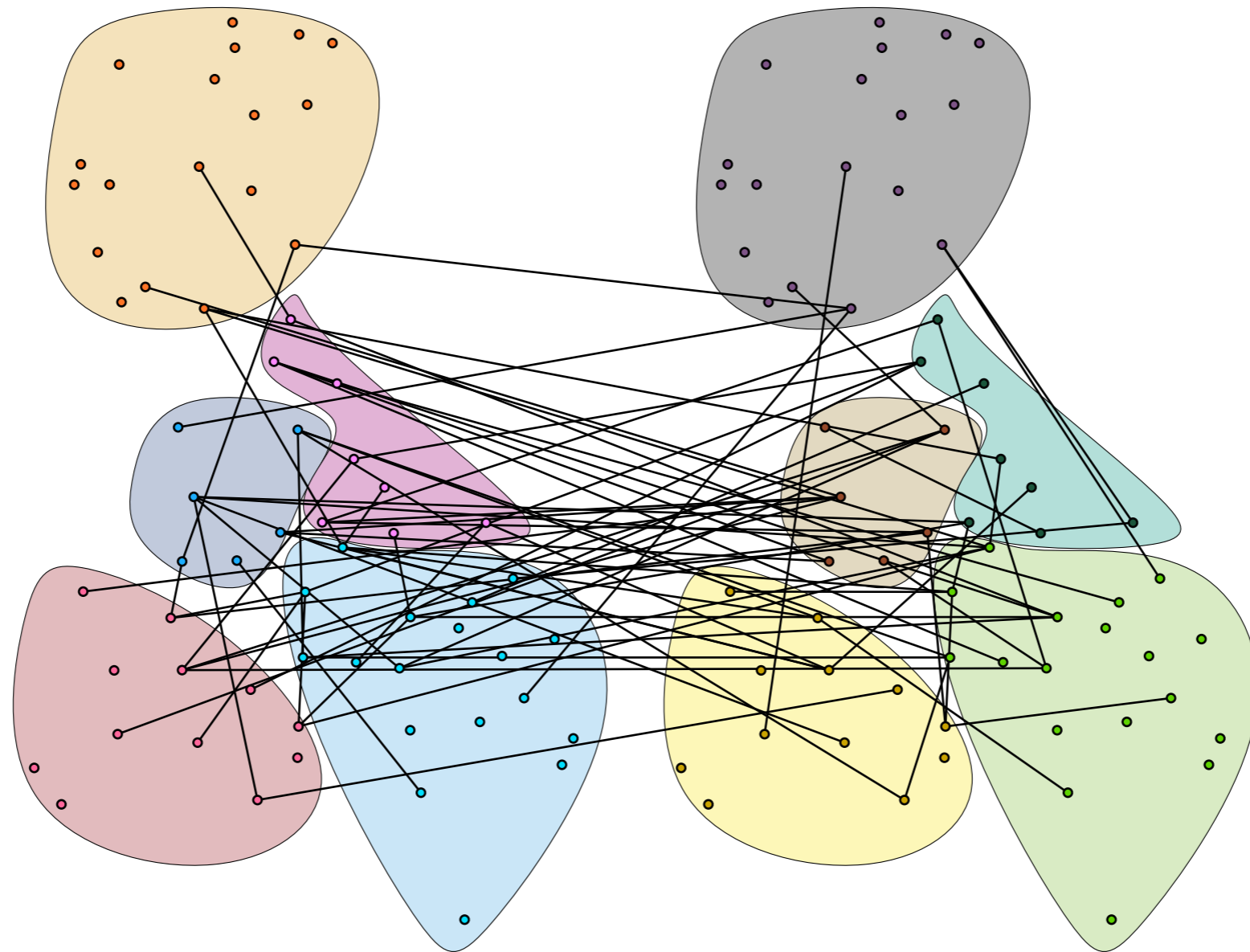
ReCoN Algorithm Example

randomized intra-community edges
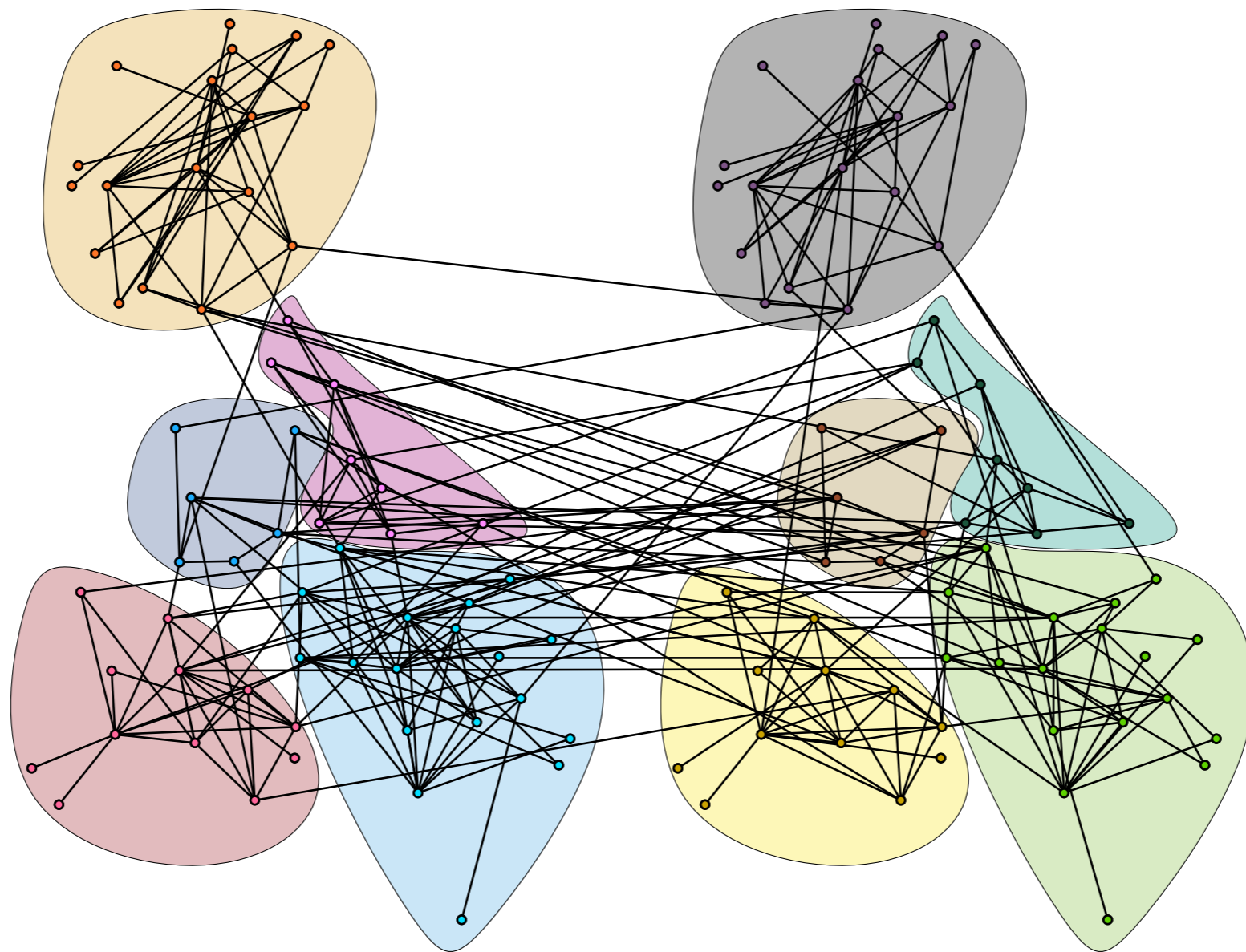
ReCoN Algorithm Example

inter-community edges

ReCoN Algorithm Example

randomized inter-community edges

ReCoN Algorithm Example

scale-2 replica

Conclusion

- **synthetic networks** are often needed in experimental algorithmics **-> generative network models**
    - **realism** is important since algorithm performance may strongly depend on structural properties
    - real data may be unavailable or unsuitable

- we show experimentally that **among a wide set of competing generative models, the ReCoN generator** is
    - best suited for creating a **randomized synthetic network** that closely replicates structural properties of the original network
    - best suited for creating a **realistically scaled-up replica** of an input network
    - **scalable** to millions of edges with our fast implementation

ReCoN Implementation

- implementation based on **NetworKit**, an open-source tool suite for the analysis of large networks
  [Staudt, Sazonovs, Meyerhenke '16 in Network Science, to appear]

Get started at https://networkit.iti.kit.edu
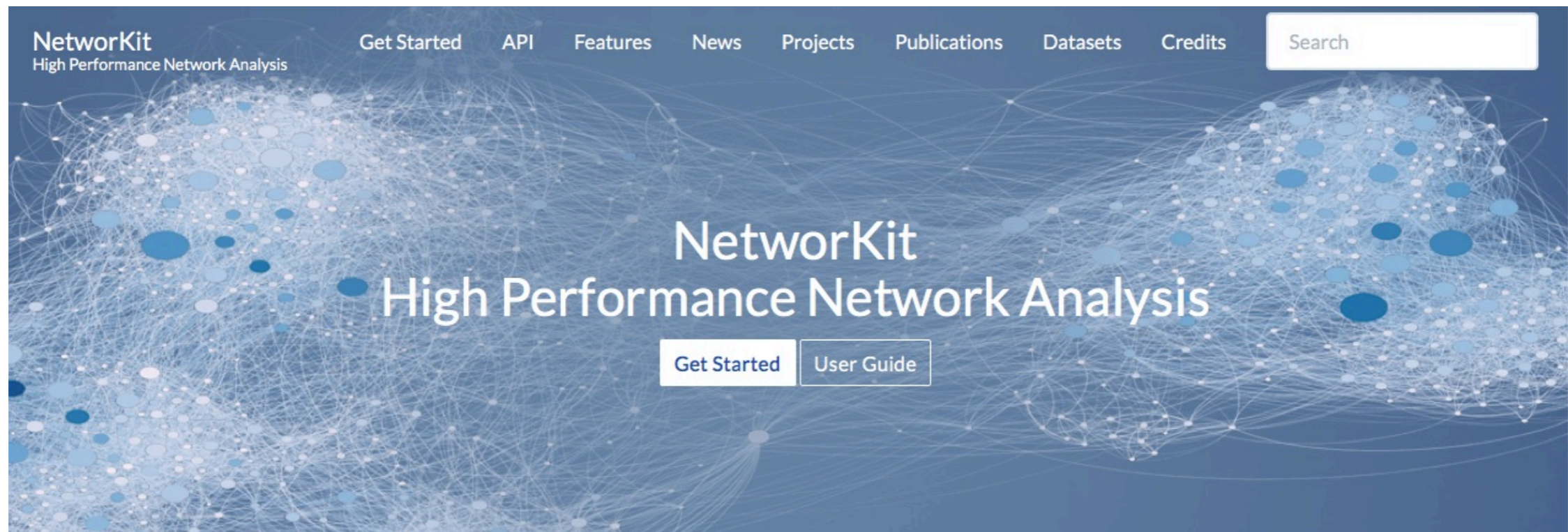
Generating Scaled Replicas of Real-World Complex Networks

Credits

CHRISTIAN L. STAUDT	data science consultant	clstaudt.me
MICHAEL HAMANN	Karlsruhe Institute of Technology	i11www.iti.kit.edu
ILYA SAFRO	Clemson University	people.cs.clemson.edu/~isafro/
ALEXANDER GUTFRAIND	University of Illinois / Uptake	gutfraind.com
HENNING MEYERHENKE	Karlsruhe Institute of Technology	parco.iti.kit.edu