icelab INTEGRATED SCIENCE LAB

UMEÅ UNIVERSITY

Mapping nonlocal relationships between metadata and network structure with metadata-dependent encoding of random walks

Martin Rosvall

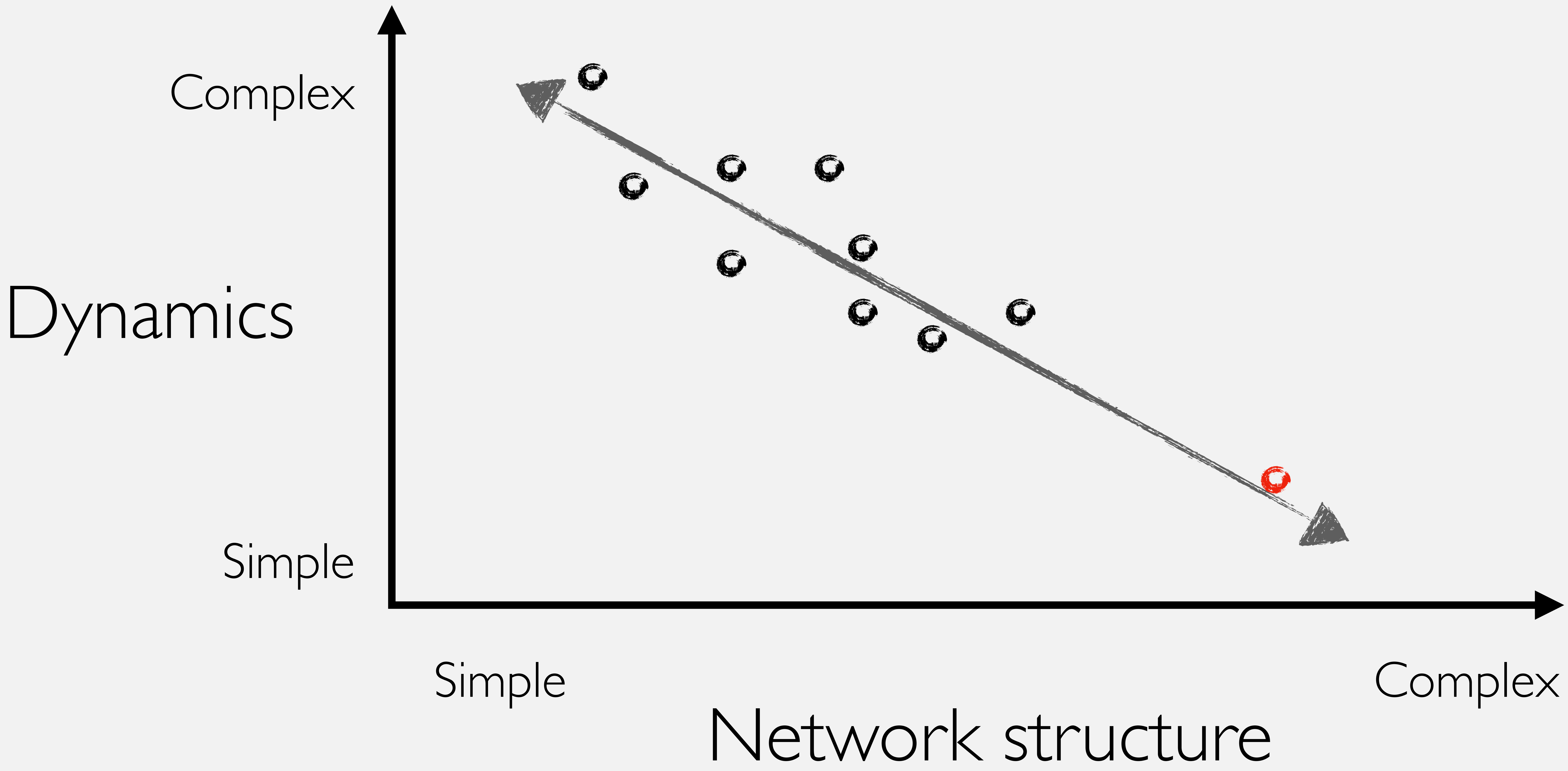
Anton Holmgren

Vincenzo Nicosia

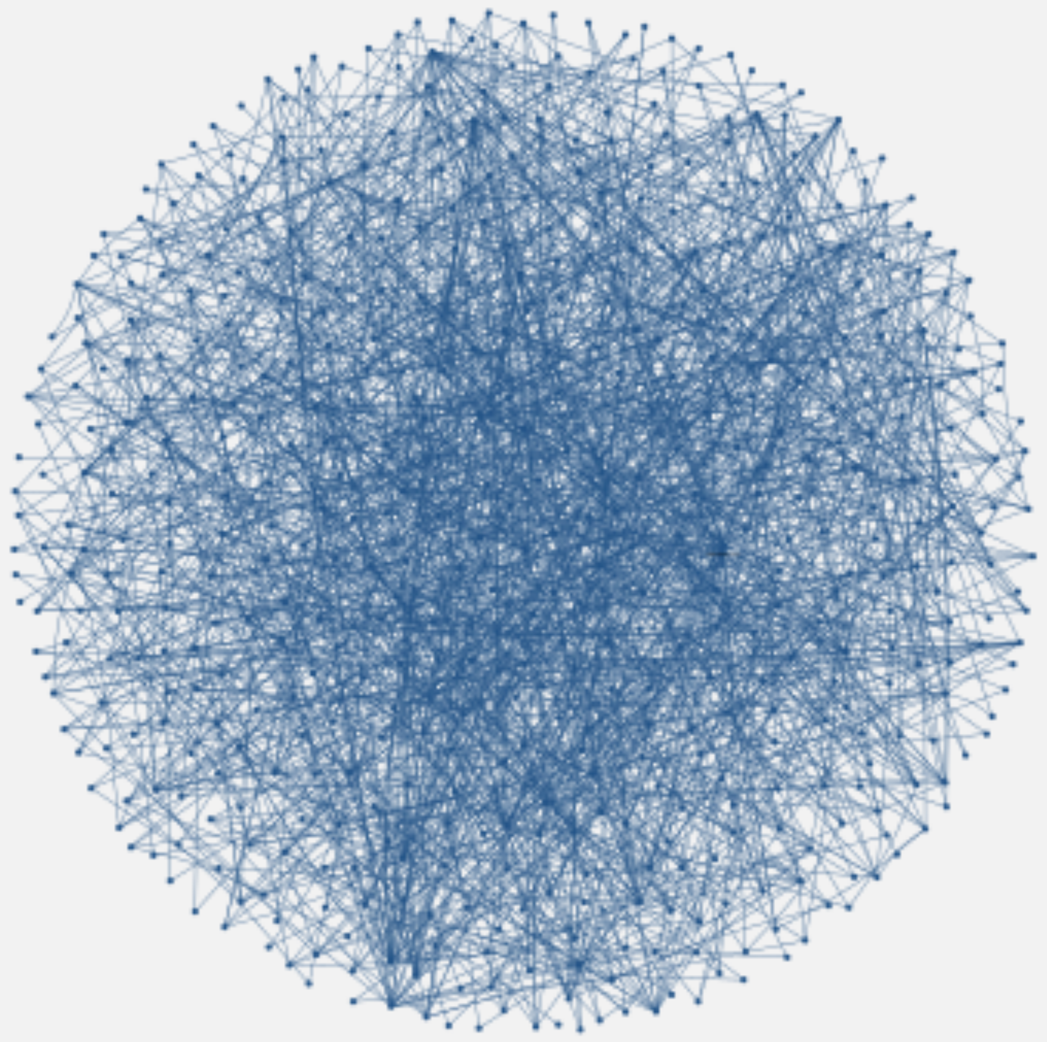
Aleix Bassolas

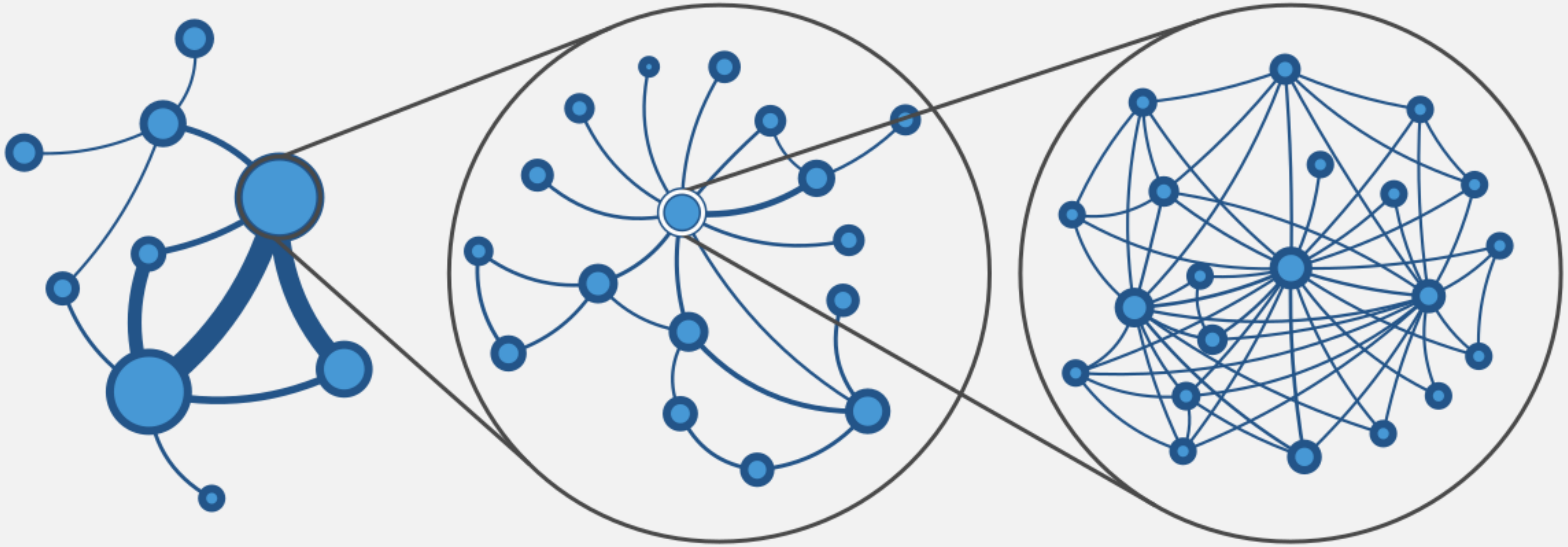
Antoine Marot

ELLIIT Focus Period 2023 on Network Dynamics and Control



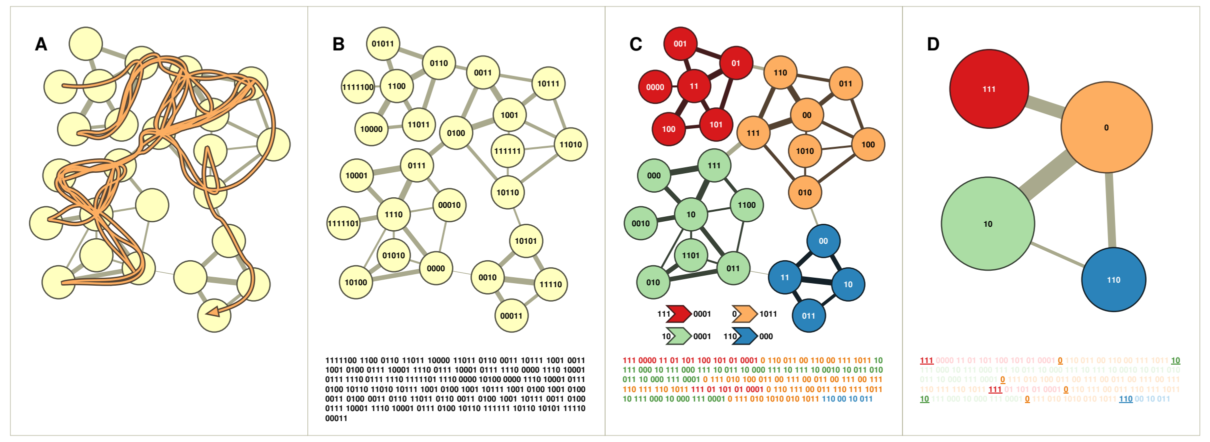
Mapping network structure

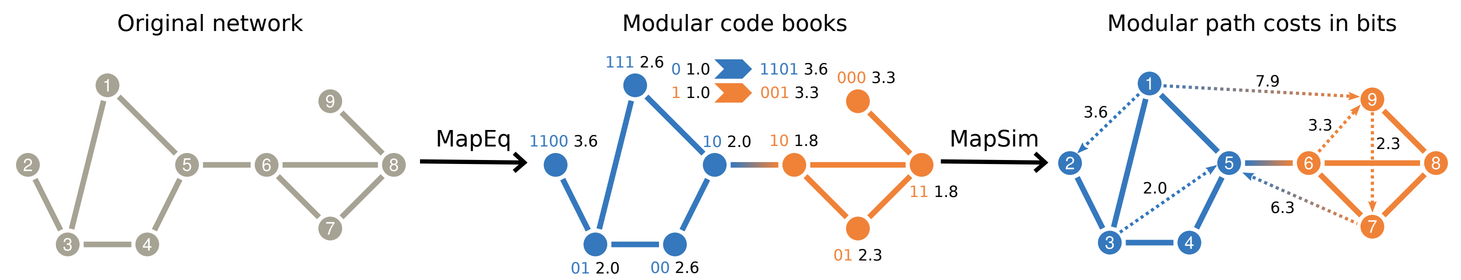




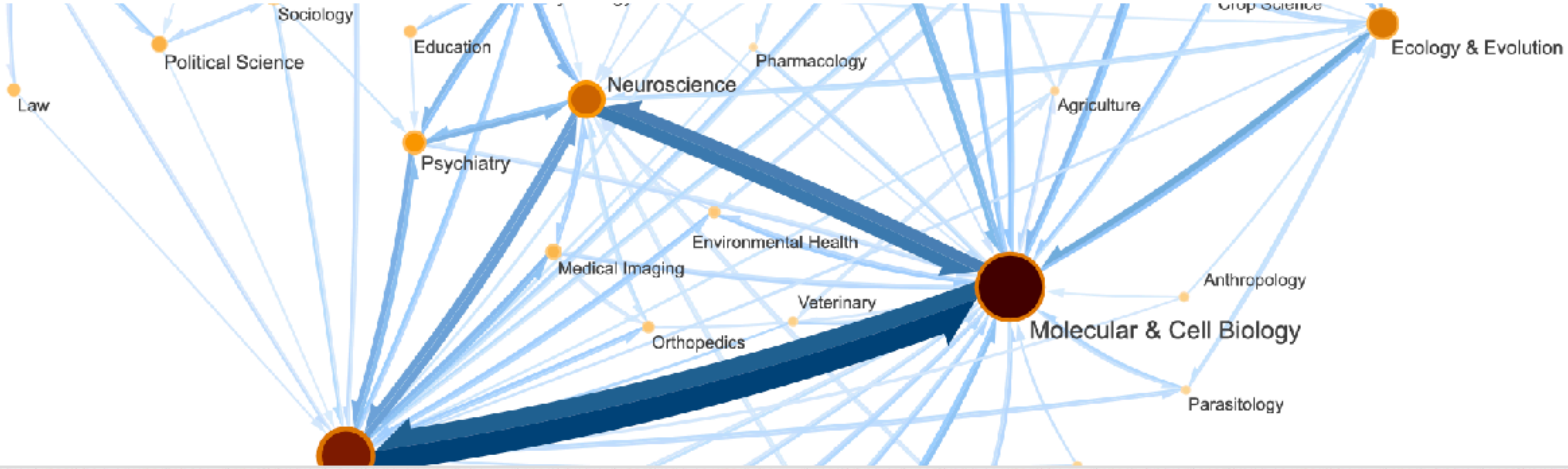
Mapping network structure

	Function 1	Function 2
	Exploration	**Navigation**
Geographic Maps	Simplify and highlight	Show directions
Google Maps for networks	Organization Function Understanding	Recommendations Drug repurposing Precision medicine
	Network theory + information theory	Network theory + information theory + machine learning

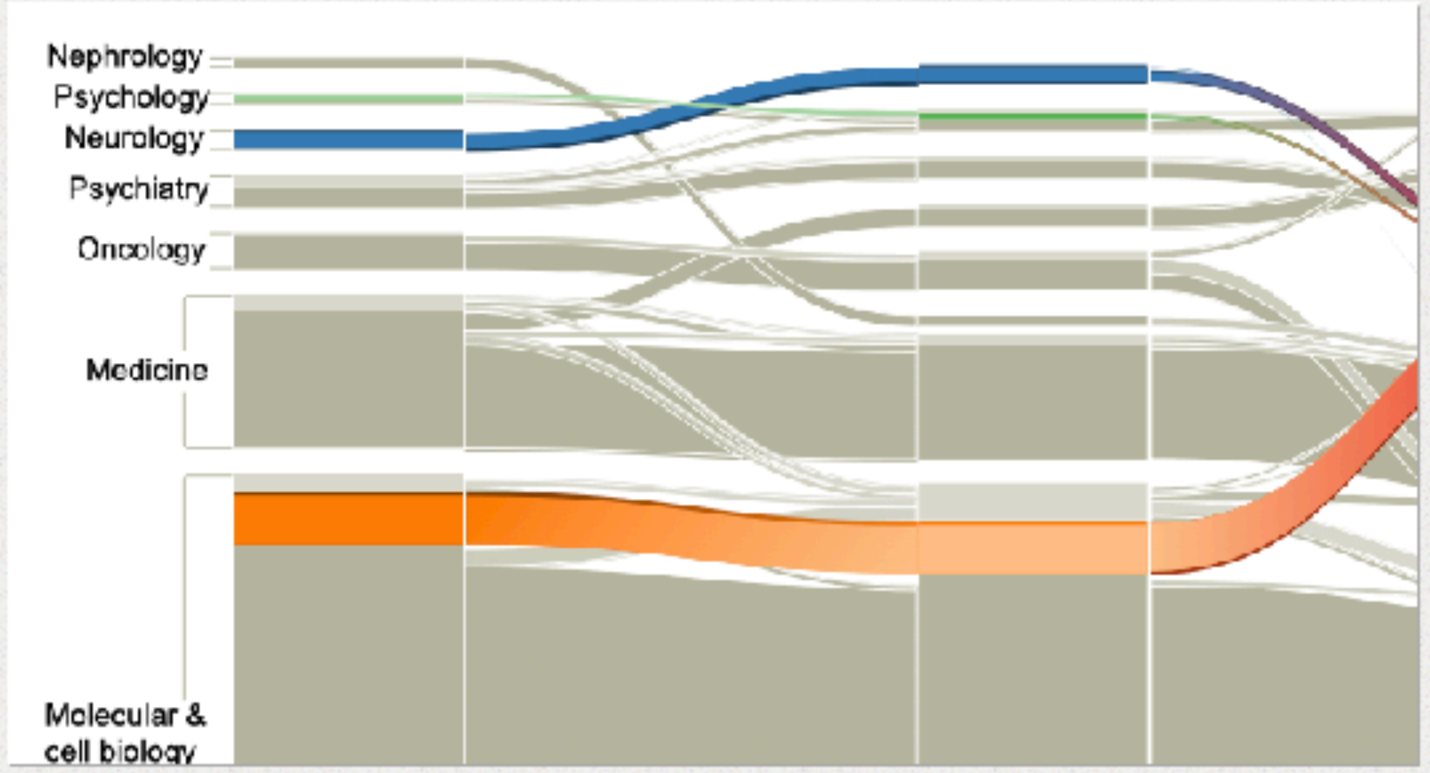




Simplify and highlight important structures in complex networks



Apps »



Code »

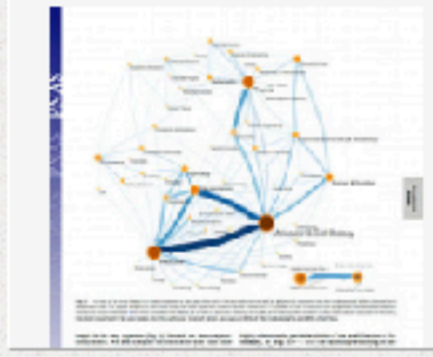
```python
from infomap import Infomap
im = Infomap()
im.read_file("ninetriangles.net")
im.add_link(1, 10)
im.run("--two-level --num-trials 5")
print(im.codelength)
for node in im.tree:
  if node.is_leaf:
    print(node.node_id, node.module_id)
```

Publications »

Maps of information flow reveal community structure in complex networks

Martin Rosvall and Carl T. Bergstrom

PNAS **105**, 1118 (2008). [arXiv:0707.0609]



To comprehend the multipartite organization of large-scale biological and social systems, we introduce a new information-theoretic approach to reveal community structure in

News

Jan 29, 2022 Preprint — Map Equation Centrality: A Community-Aware Centrality Score Based on the Map Equation — arXiv:2201.12590

Dec 10, 2021 Research paper — Mapping flows on weighted and directed networks with incomplete observations — J. Comp. Net. 9, 6 (2021)

Mapping network structure and metadata on power grids

"We would like to play with some parameters to give more importance to some nodes in the graph, somehow conditioning the clustering when we are interested in specific elements of the graph.

Antoine Marot

Lead AI Scientist at RTE – France's Transmission System Operator

Mapping network structure and metadata on power grids

"We want coherent communities with nodes that share similar prices.

Antoine Marot

Lead AI Scientist at RTE – France's Transmission System Operator

Network structure and metadata

Question: How can we exploit nonlocal relationships between network structure and metadata?

nature COMMUNICATIONS

ARTICLE

Received 12 Feb 2016 | Accepted 5 May 2016 | Published 16 Jun 2016 DOI: 10.1038/ncomms11863 OPEN

Structure and inference in annotated networks

M.E.J. Newman[1,2,3] & Aaron Clauset[3,4,5]

For many networks of scien
and information about the
social network. Here we de
understanding of network st
detection in networks and
network and its metadata to detect communities more accurately than can be done with either alone. Crucially, the method does not assume that the metadata are correlated with the communities we are trying to find. Instead, the method learns whether a correlation exists and correctly uses or ignores the metadata depending on whether they contain useful information. We demonstrate our method on synthetic networks with known structure and on real-world networks, large and small, drawn from social, biological and technological domains.

Require correlations between network structure and metadata

PHYSICAL REVIEW E **100**, 022301 (2019)

Map equation with metadata: Varying the role of attributes in community detection

Scott Emmons* and Peter J. Mucha†
Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics, University of North Carolina, Chapel Hill, North Carolina 27599, USA

(Received 24 October 2018; revised manuscript received 17 July 2019; published 2 August 2019)

Much of the community detection literature studies structural communities, communities defined solely by the connectivity patterns of the network. Often networks contain additional metadata which can inform community detection such as the grade and gender of students in a high school social network. In this work, we introduce a
ity detection algorithm to
ic networks, we show that
l aligned with community
tual information with the
ng knob of a microscope,
s on the metadata.

I. INTRODUCTION

As network science has found application in a variety of real-world systems, ranging from the biological to the technological, so, too, has community detection in networks received widespread attention [1–4]. Traditionally, community detection methods have focused solely on the topology of the network, optimizing an objective function defined on the network structure that captures a particular notion of community, such as intracommunity edge density and inter-

maximization by including attributes in their "intertia-based modularity." Yang *et al.* proposed CESNA [12] and He *et al.* proposed CNMMA [13] to identify communities by learning a latent space that generates links and attributes. Peel *et al.* [6] established a statistical test to determine if attributes correlate with community structure, and they developed an SBM with flexibility in how strongly to couple attributes and community labels in the corresponding stochastic block model inference. In related work, Stanley *et al.* [14] propose a test statistic based on label propagation for the alignment of node attributes

1. Mapping network flows

2. with metadata-dependent encoding

3. exploits nonlocal relationships

1.

Mapping network flows

using the minimum description length principle

MAPS depict regularities
using less information

NETWORKS describe where flows move
depending on the current node

Mapping network flows
using the minimum description length principle

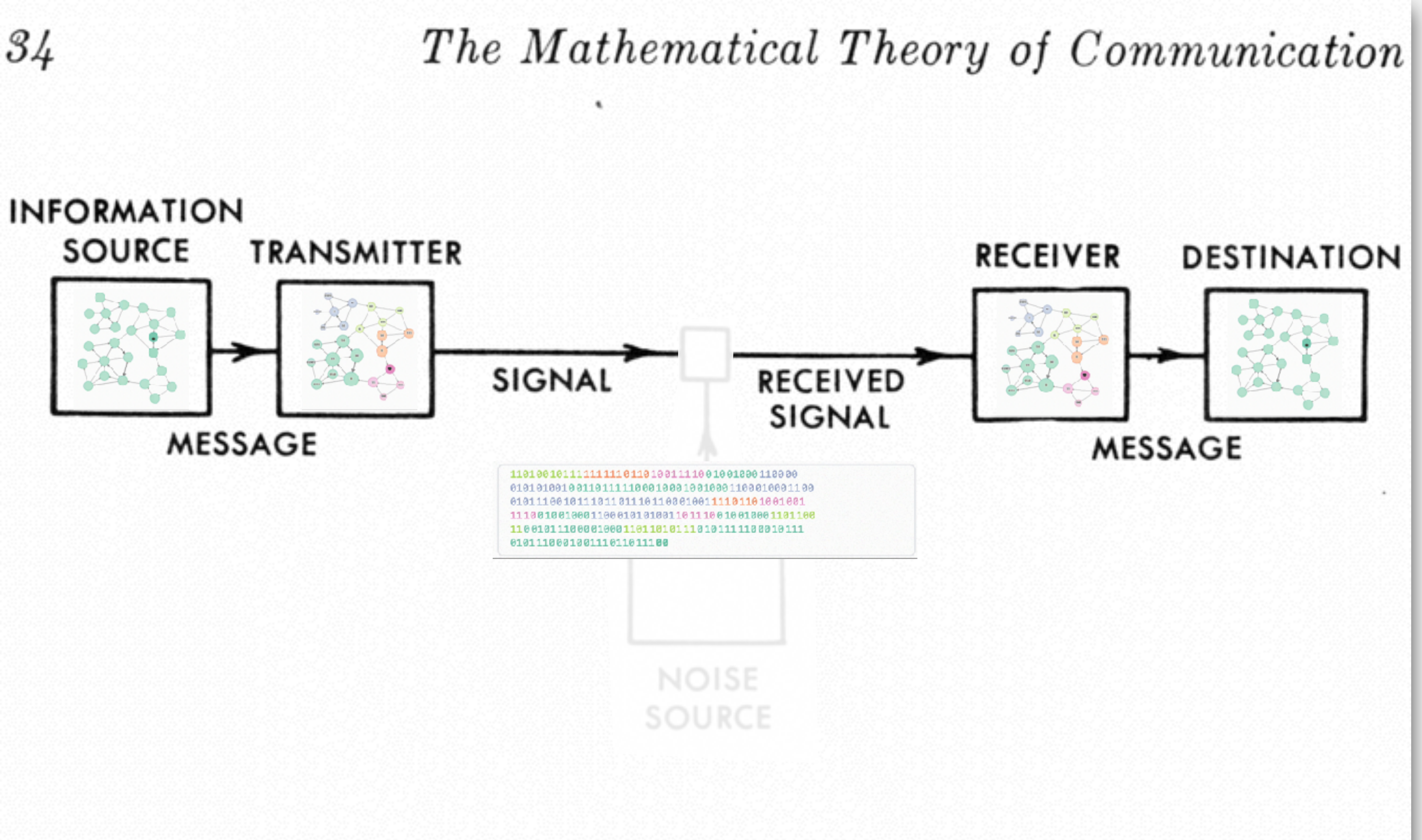


Fig. 1. — Schematic diagram of a general communication system.

Mapping network flows

using the minimum description length principle

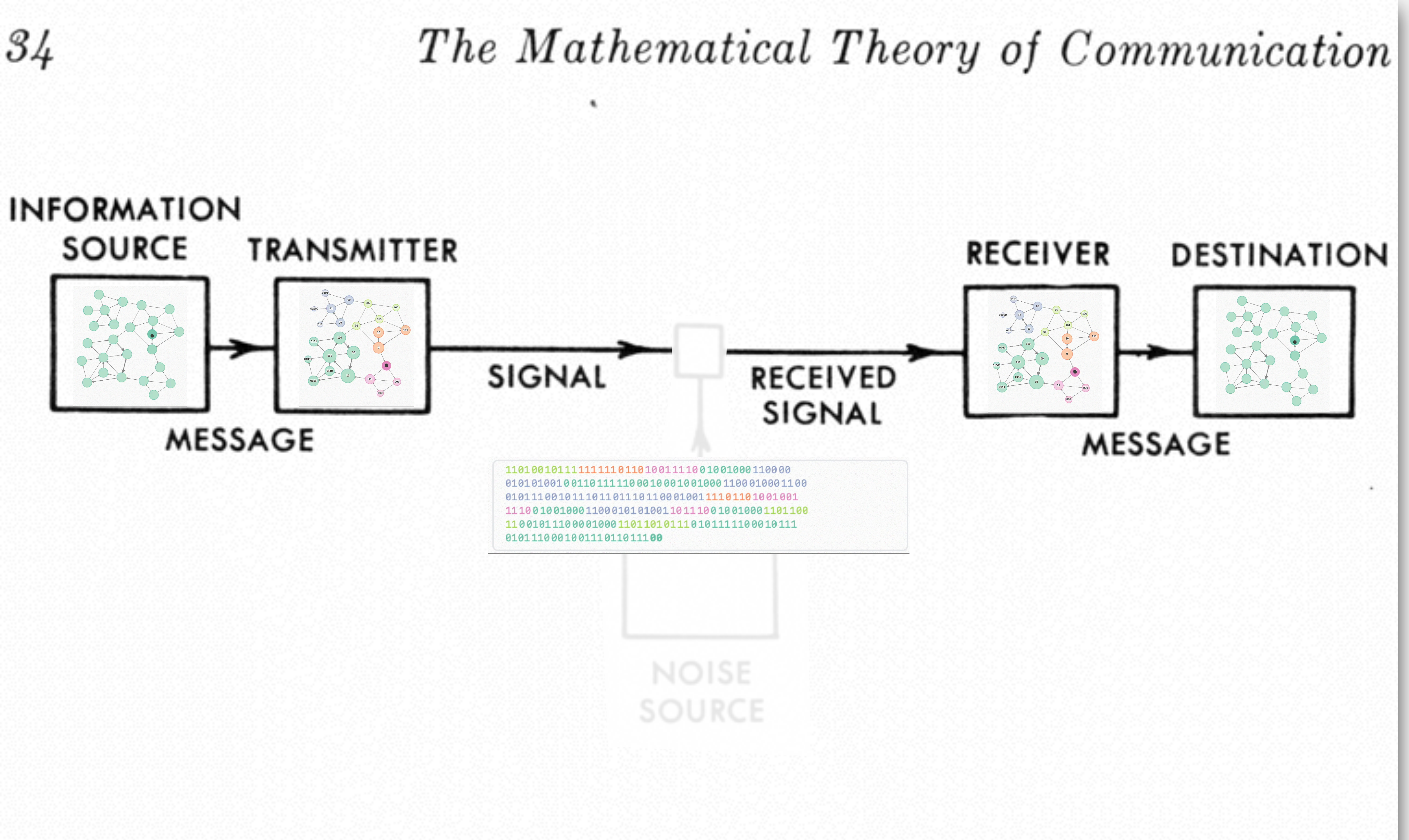
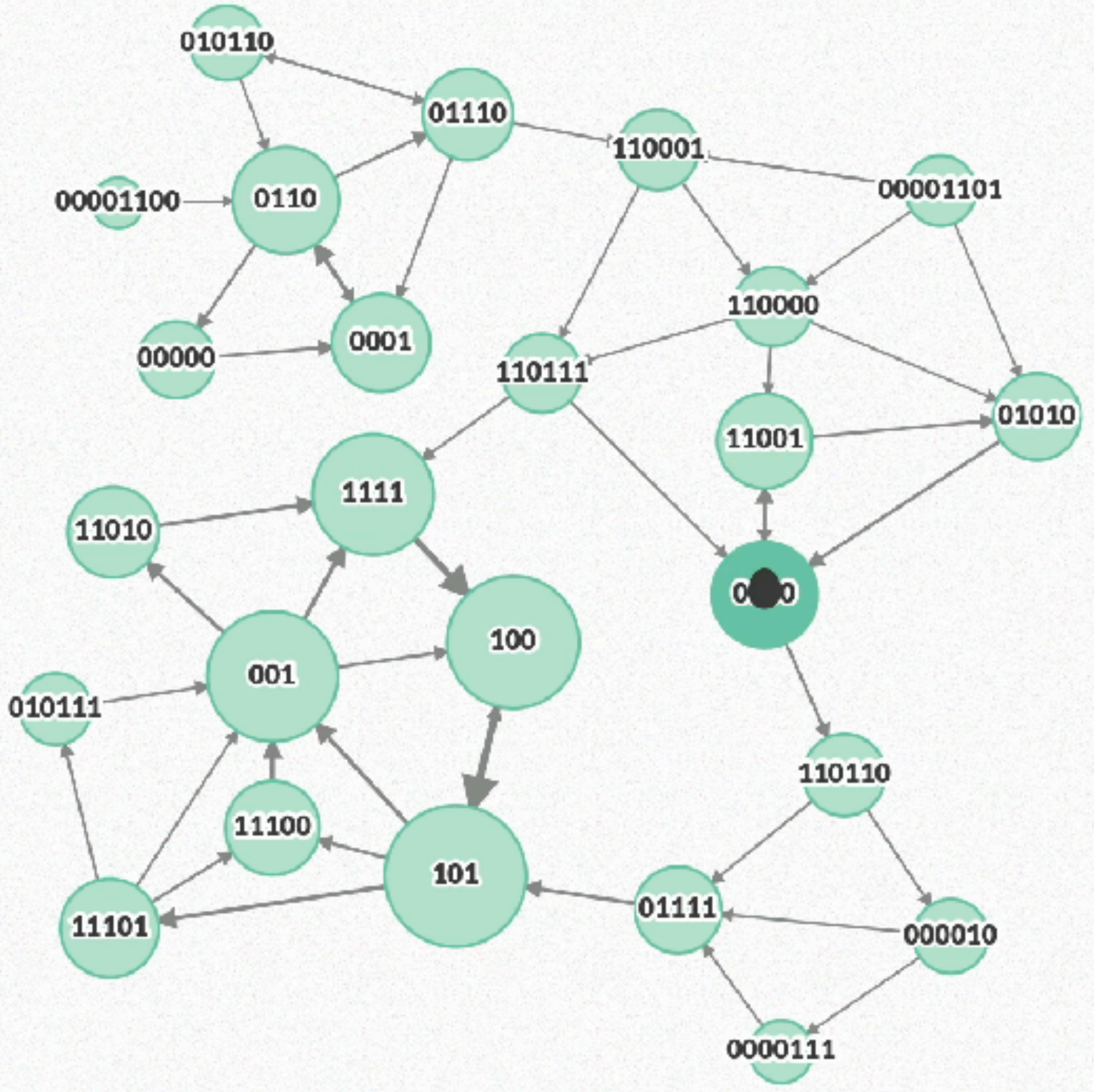
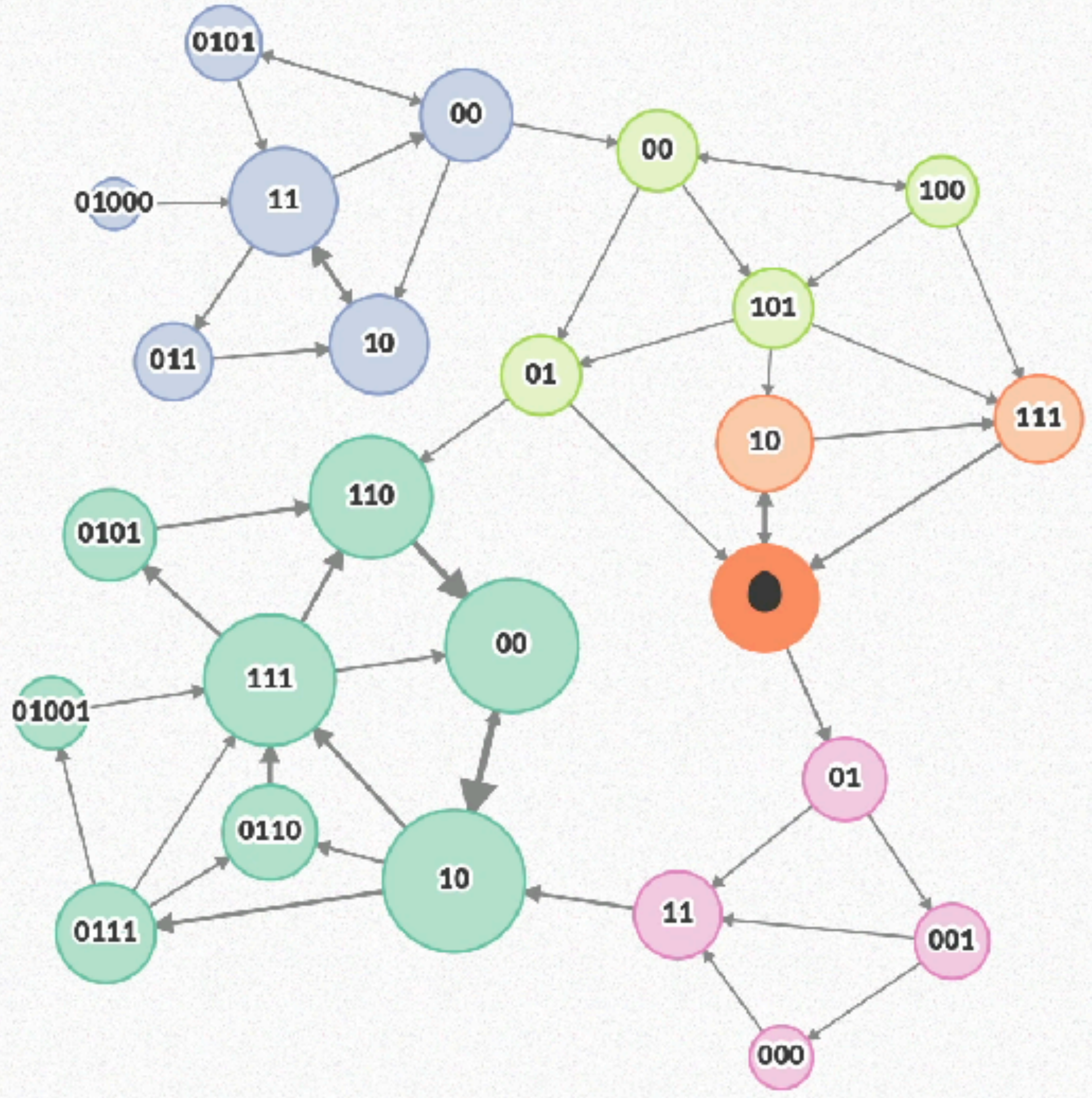


Fig. 1. — Schematic diagram of a general communication system.

Mapping network flows

using the minimum description length principle






```
00000 0001 0110 0001 0110 0001 0110 00000 0001 11001 01010
0100 11001 01010 0100 11001 01010 0100 11001 0100 11001 0100
11011 0 01111 101 001 100 101 11101 001 11010 1111 100 101 001
100 101 11100 001 11010 1111 11001 0100 11001 0100 110110
01111 101 11100 001 1111 01010 0100 11001 01010 0100 11001
0100
```

Average codelength: 4.321 bits

```
1011100111011011111000100111111001011111000
010001100011101110111011011100100111110111010
11101011101001001101001111001011100100111
1110101110001011100100110111010111001000111100
10011010011110010011011111001000111111010111
0100
```

Average codelength: 3.765 bits

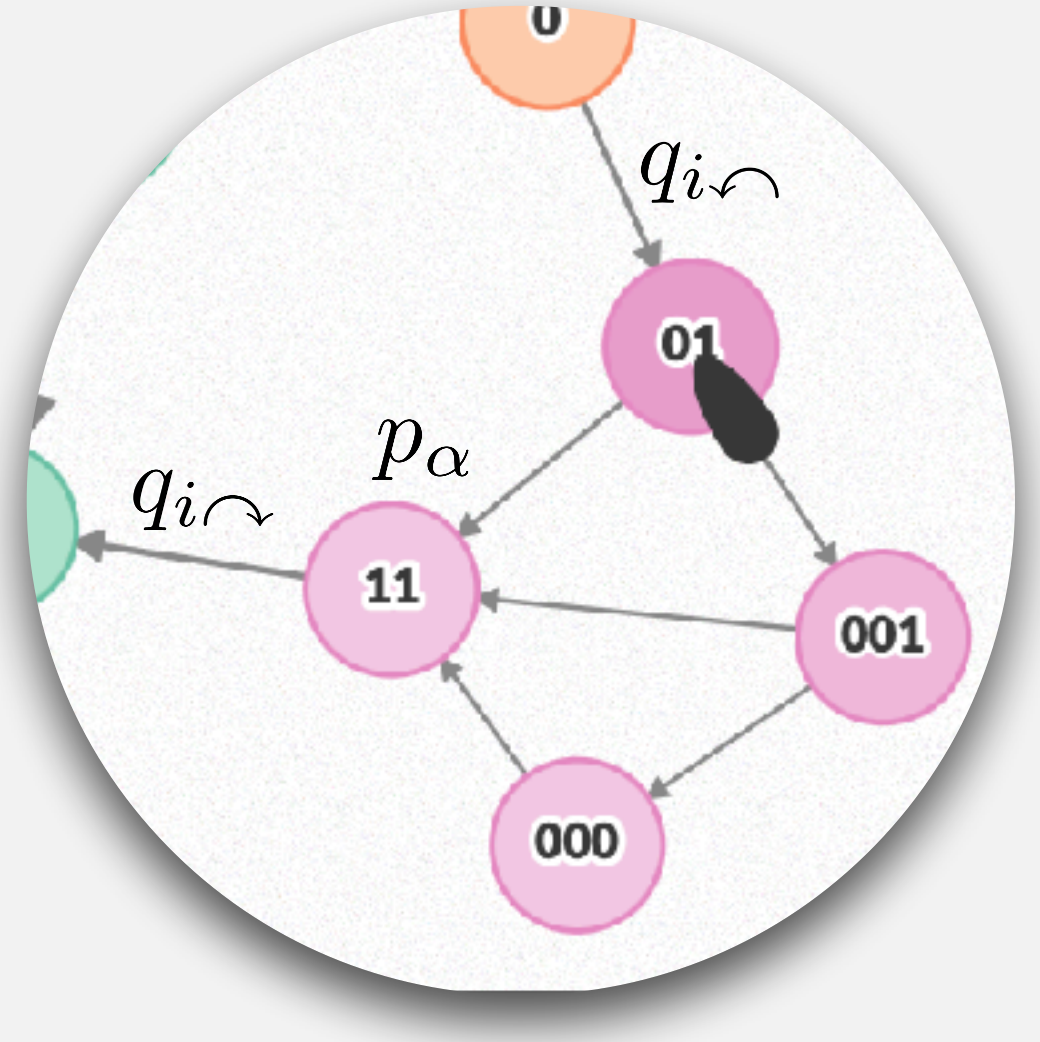
$$L(\mathsf{M}_1) = H(\mathcal{P}) = 4.23 \text{ bits}$$

$$L(\mathsf{M}_5) = q_{\curvearrowleft} H(\mathcal{Q}) + \sum_{i=1}^{5} p_i^{\circlearrowright} H(\mathcal{P}_i) = 0.42 + 3.13 = 3.55 \text{ bits}$$

Mapping network flows
using the map equation

Visit and transition rates

$$\boldsymbol{\rho} = (p_\alpha, q_{i\curvearrowleft}, q_{i\curvearrowright})$$



Per-step use rate of the index codebook.

$$q_{\curvearrowleft} = \sum_{i=1}^{m} q_{i\curvearrowleft}$$

Per-step use rate of module codebook i

$$p_i^{\circlearrowright} = q_{i\curvearrowright} + \sum_{\alpha \in \mathsf{M}_i} p_\alpha$$

$$L(\mathsf{M}) = q_{\curvearrowleft} H(\mathcal{Q}) + \sum_{i=1}^{m} p_i^{\circlearrowright} H(\mathcal{P}_i)$$

$$\mathcal{Q} = \{q_{i\curvearrowleft}\}$$

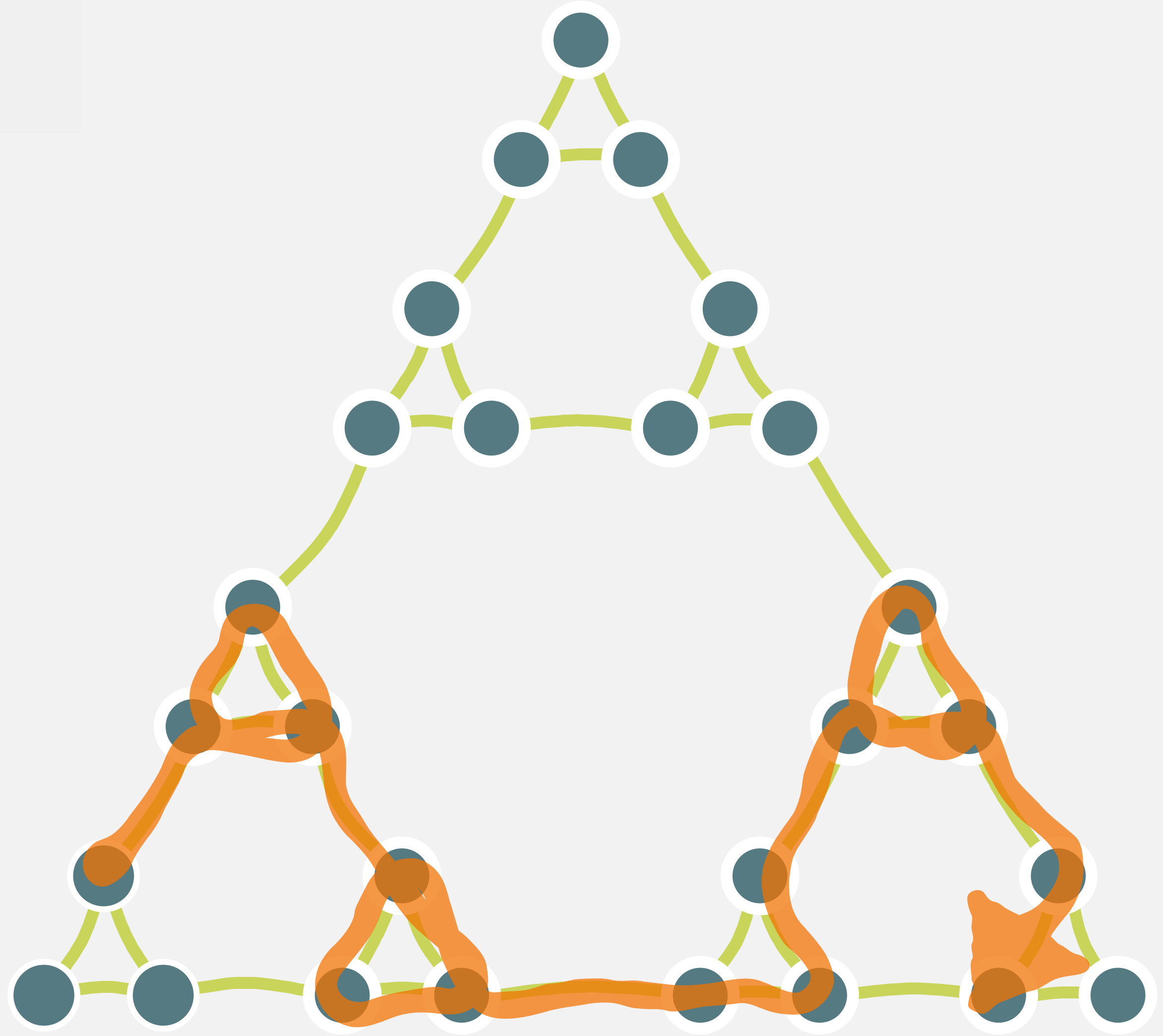
Per-step average code length of index codebook for steps between modules

$$\mathcal{P}_i = \{q_{i\curvearrowright}, p_{\alpha \in \mathsf{M}_i}\}$$

Per-step average code length of module codebook i for steps in module i

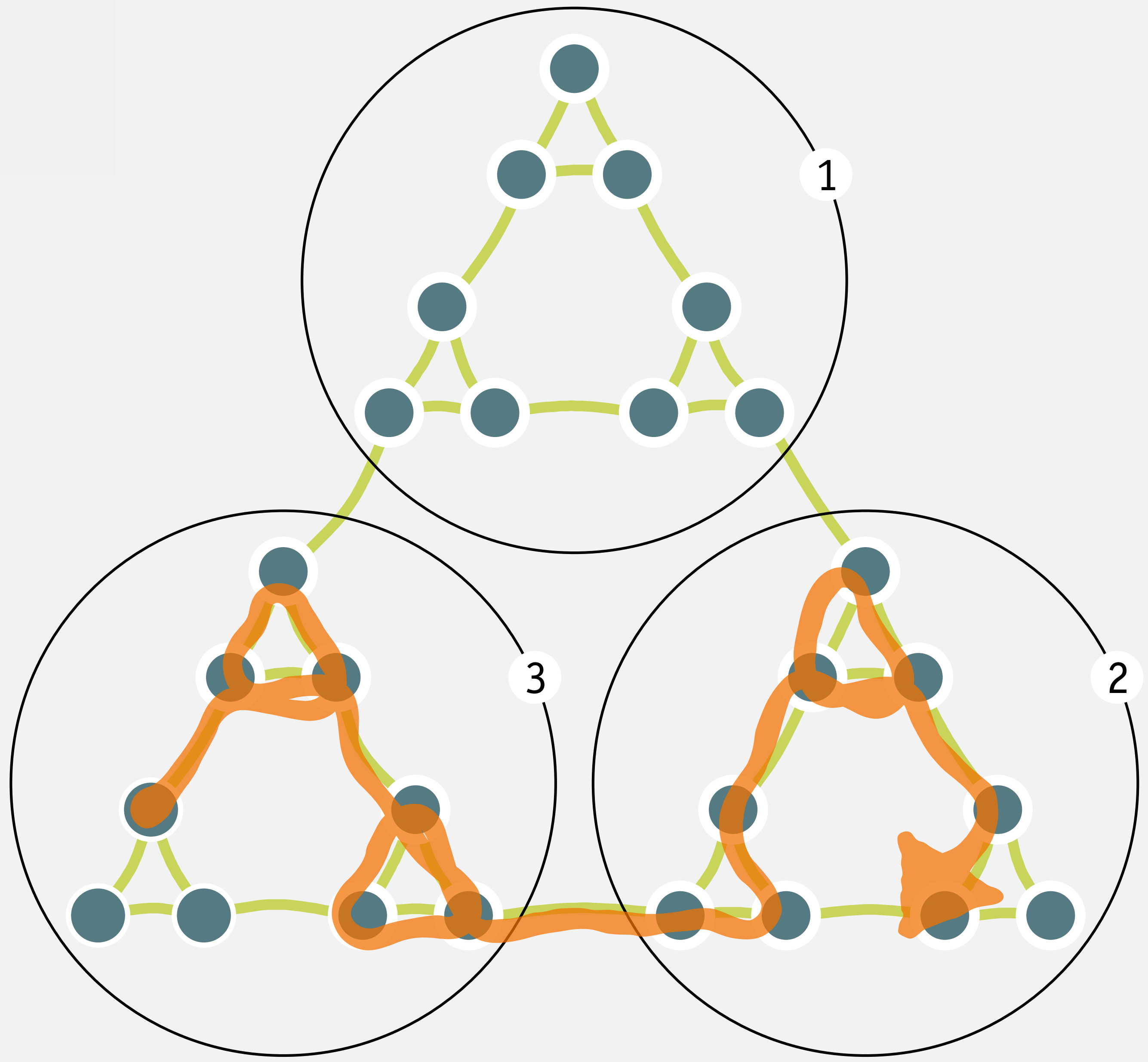
The map equation

Compression of network flows



21

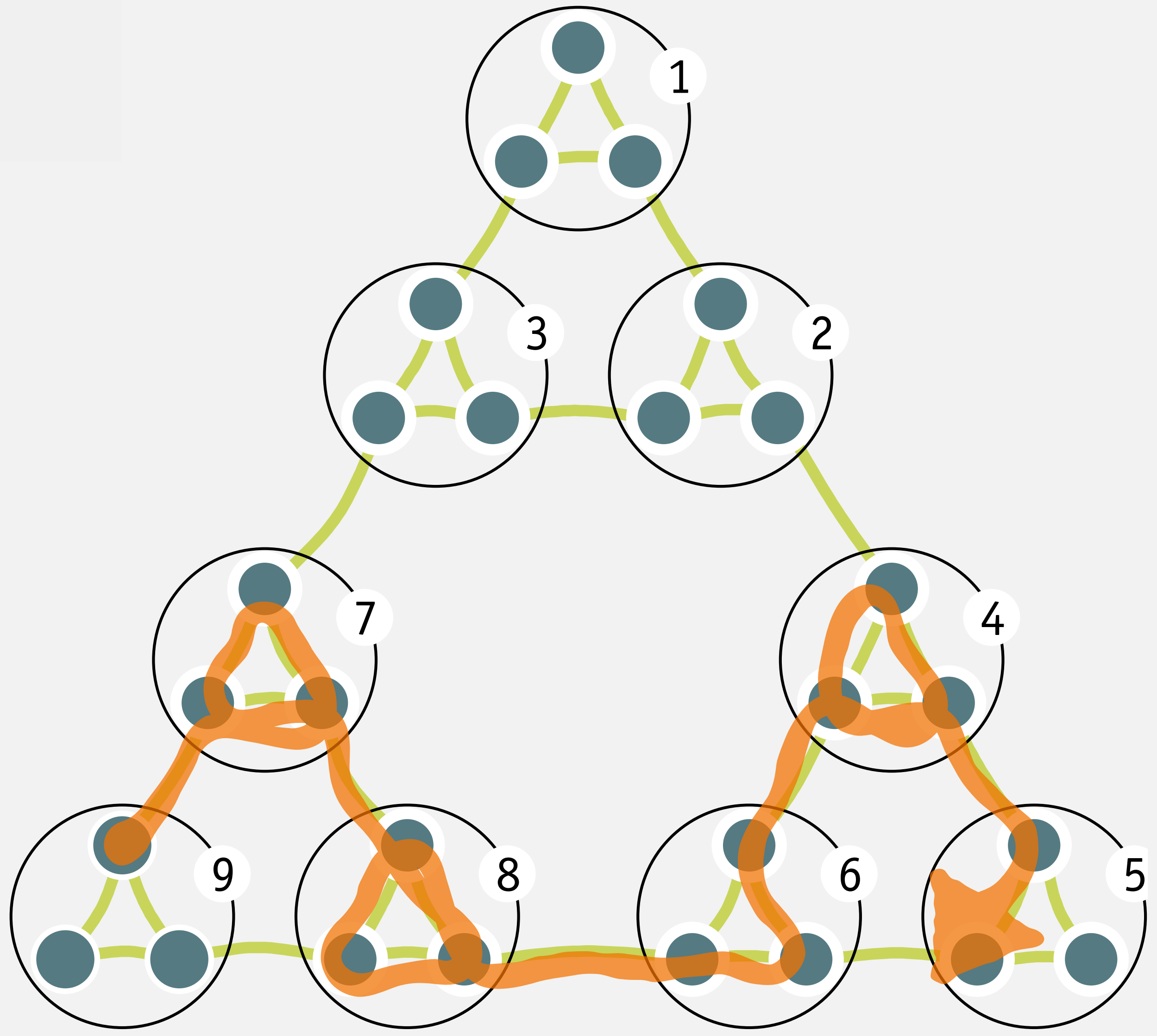
$$L(\mathsf{M}) = H(\mathcal{P}) = 4.75 \text{ bits.}$$



1

22

$$L(\mathsf{M}) = \underbrace{q_{\curvearrowright} H(\mathcal{Q})}_{0.12 \text{ bits}} + \underbrace{\begin{cases} p^1_{\circlearrowright} H(\mathcal{P}^1) \\ p^2_{\circlearrowright} H(\mathcal{P}^2) \\ p^3_{\circlearrowright} H(\mathcal{P}^3) \end{cases}}_{3.56 \text{ bits}} = 3.68 \text{ bits.}$$



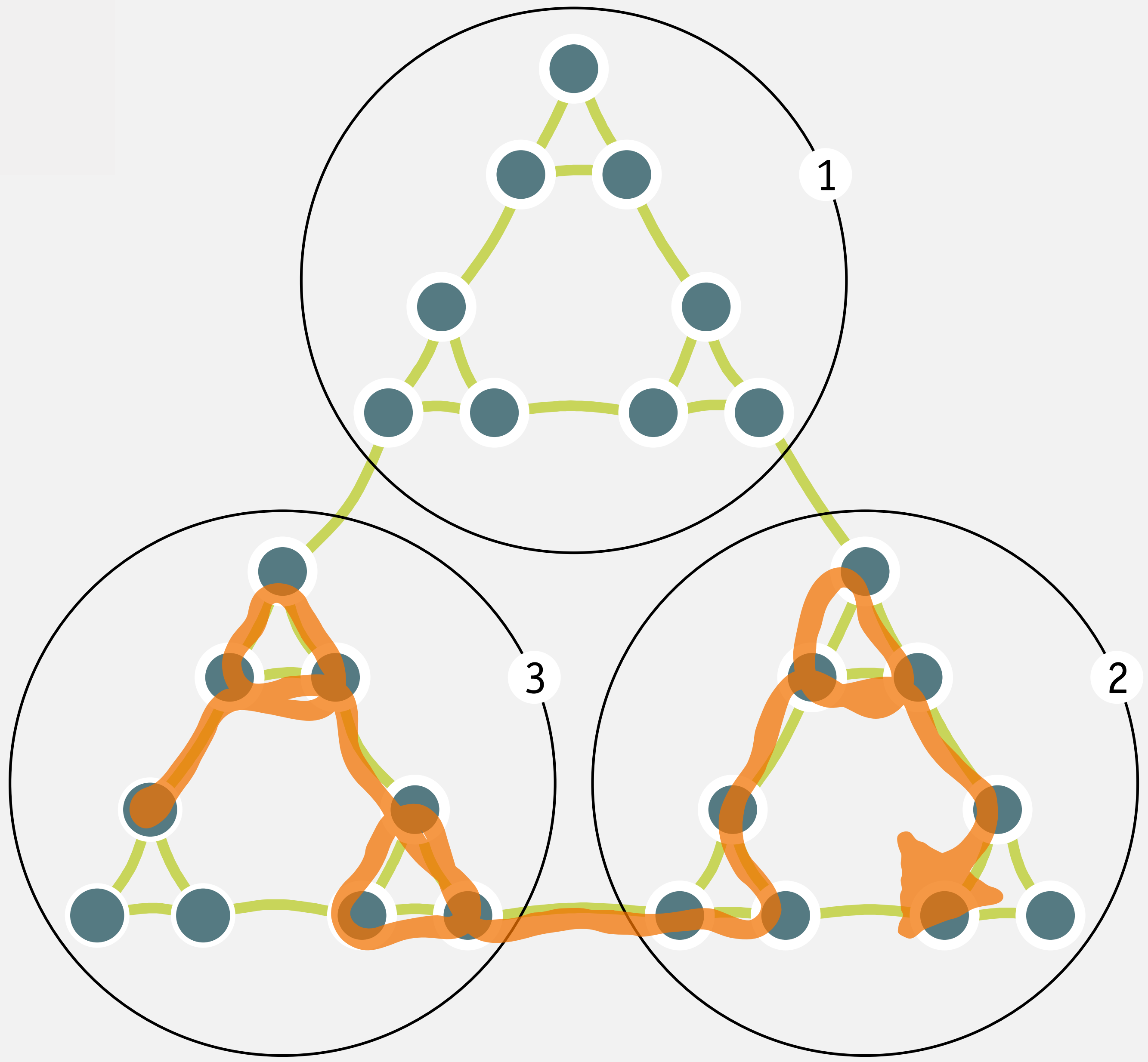
5

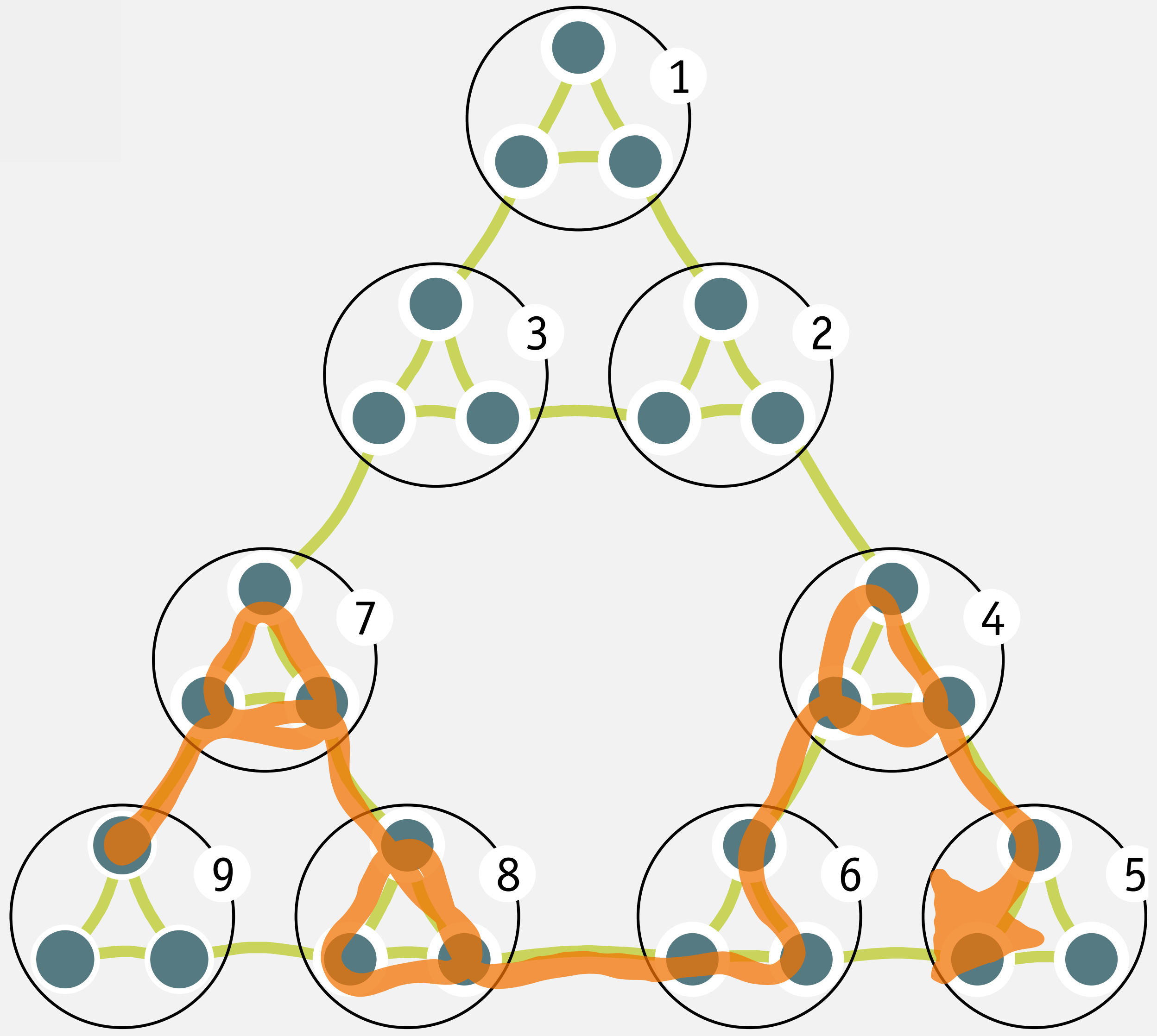
26

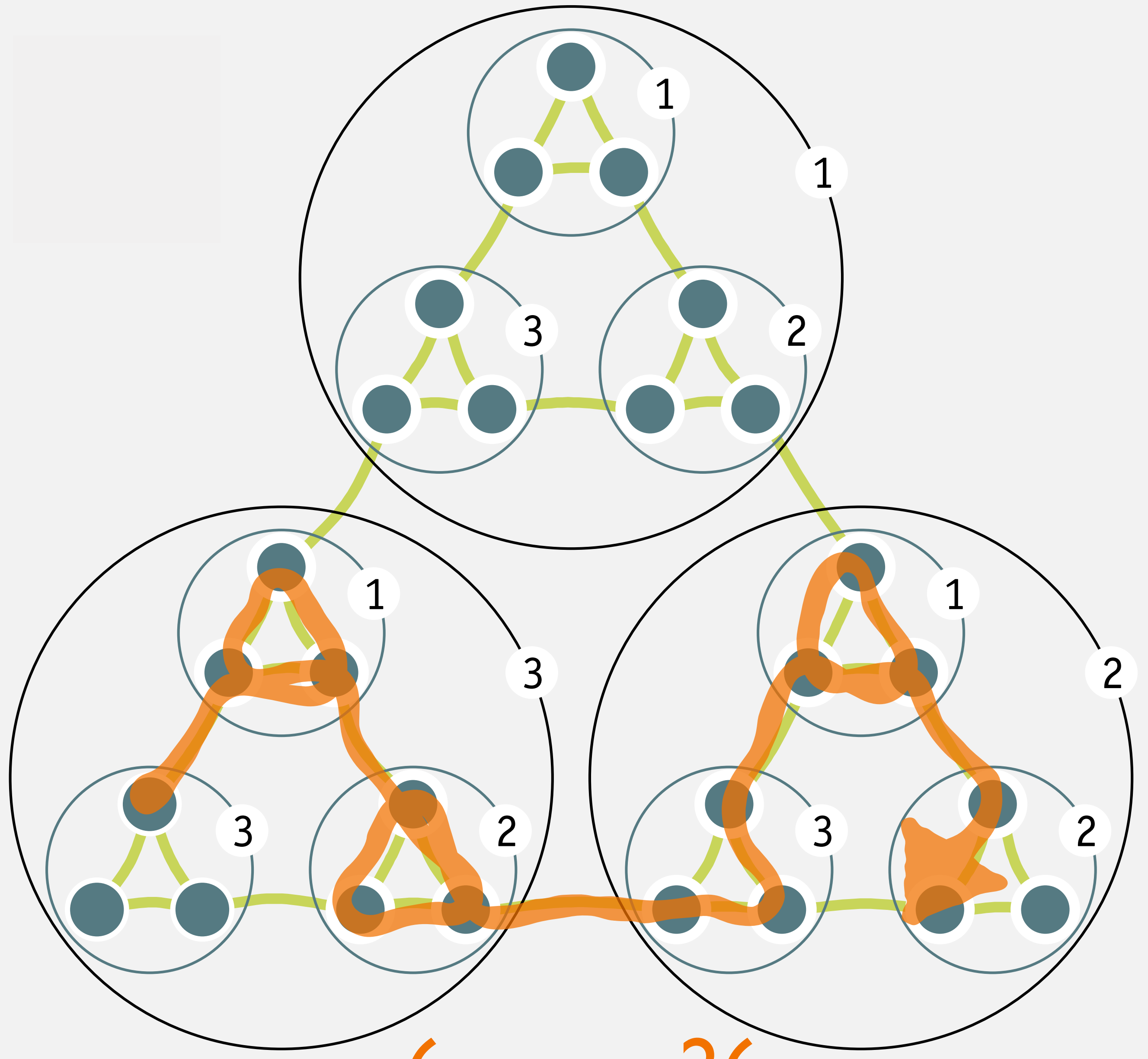
$$L(\mathsf{M}) = \underbrace{q_{\curvearrowright} H(\mathcal{Q})}_{0.97 \text{ bits}} + \underbrace{\begin{cases} p^1_{\circlearrowright} H(\mathcal{P}^1) \\ p^2_{\circlearrowright} H(\mathcal{P}^2) \\ p^3_{\circlearrowright} H(\mathcal{P}^3) \\ p^4_{\circlearrowright} H(\mathcal{P}^4) \\ p^5_{\circlearrowright} H(\mathcal{P}^5) \\ p^6_{\circlearrowright} H(\mathcal{P}^6) \\ p^7_{\circlearrowright} H(\mathcal{P}^7) \\ p^8_{\circlearrowright} H(\mathcal{P}^8) \\ p^9_{\circlearrowright} H(\mathcal{P}^9) \end{cases}}_{2.60 \text{ bits}} = 3.57 \text{ bits.}$$

The map equation

Compression of network flows







1 22

$$L(\mathsf{M}) = \underbrace{q_{\curvearrowleft} H(\mathcal{Q})}_{0.12 \text{ bits}} + \underbrace{\left\{ \begin{array}{l} p^1_{\circlearrowright} H(\mathcal{P}^1) \\ p^2_{\circlearrowright} H(\mathcal{P}^2) \\ p^3_{\circlearrowright} H(\mathcal{P}^3) \end{array} \right.}_{3.56 \text{ bits}} = 3.68 \text{ bits.}$$

5 26

$$L(\mathsf{M}) = \underbrace{q_{\curvearrowleft} H(\mathcal{Q})}_{0.97 \text{ bits}} + \underbrace{\left\{ \begin{array}{l} p^1_{\circlearrowright} H(\mathcal{P}^1) \\ p^2_{\circlearrowright} H(\mathcal{P}^2) \\ p^3_{\circlearrowright} H(\mathcal{P}^3) \\ p^4_{\circlearrowright} H(\mathcal{P}^4) \\ p^5_{\circlearrowright} H(\mathcal{P}^5) \\ p^6_{\circlearrowright} H(\mathcal{P}^6) \\ p^7_{\circlearrowright} H(\mathcal{P}^7) \\ p^8_{\circlearrowright} H(\mathcal{P}^8) \\ p^9_{\circlearrowright} H(\mathcal{P}^9) \end{array} \right.}_{2.60 \text{ bits}} = 3.57 \text{ bits.}$$

1 6 26

$$L(\mathsf{M}) = \underbrace{q_{\curvearrowleft} H(\mathcal{Q})}_{0.12 \text{ bits}} + \underbrace{\left\{ \begin{array}{l} q^1_{\circlearrowright} H(\mathcal{Q}^1) + \\ \\ q^2_{\circlearrowright} H(\mathcal{Q}^2) + \\ \\ q^3_{\circlearrowright} H(\mathcal{Q}^3) + \end{array} \right.}_{0.76 \text{ bits}} \underbrace{\begin{array}{l} \left\{ \begin{array}{l} p^{11}_{\circlearrowright} H(\mathcal{P}^{11}) \\ p^{12}_{\circlearrowright} H(\mathcal{P}^{12}) \\ p^{13}_{\circlearrowright} H(\mathcal{P}^{13}) \end{array} \right. \\ \left\{ \begin{array}{l} p^{21}_{\circlearrowright} H(\mathcal{P}^{21}) \\ p^{22}_{\circlearrowright} H(\mathcal{P}^{22}) \\ p^{23}_{\circlearrowright} H(\mathcal{P}^{23}) \end{array} \right. \\ \left\{ \begin{array}{l} p^{31}_{\circlearrowright} H(\mathcal{P}^{31}) \\ p^{32}_{\circlearrowright} H(\mathcal{P}^{32}) \\ p^{33}_{\circlearrowright} H(\mathcal{P}^{33}) \end{array} \right. \end{array}}_{2.60 \text{ bits}} = 3.48 \text{ bits.}$$

1. Mapping network flows

using the minimum description length principle

The map equation infers modules with long flow persistence using the minimum description length principle.

Generalizations to many network representations.

2. Mapping network flows

with metadata-dependent encoding

Mapping network flows
Absorbing random walks

1. We use random walks that remember their origin

2. Each node i has associated metadata f_i

3. The probability x_{ij} of a walker starting at i to be absorbed at j depends on f_i and f_j

Mapping network flows

Encoding probabilities for categorical metadata

Baseline encoding probability, $[\mathbf{0}, \mathbf{1}]$

Encoding probability

$$\varepsilon_{ij} = p\delta_{f_i, f_j} + \frac{p}{c}(1 - \delta_{f_i, f_j})$$

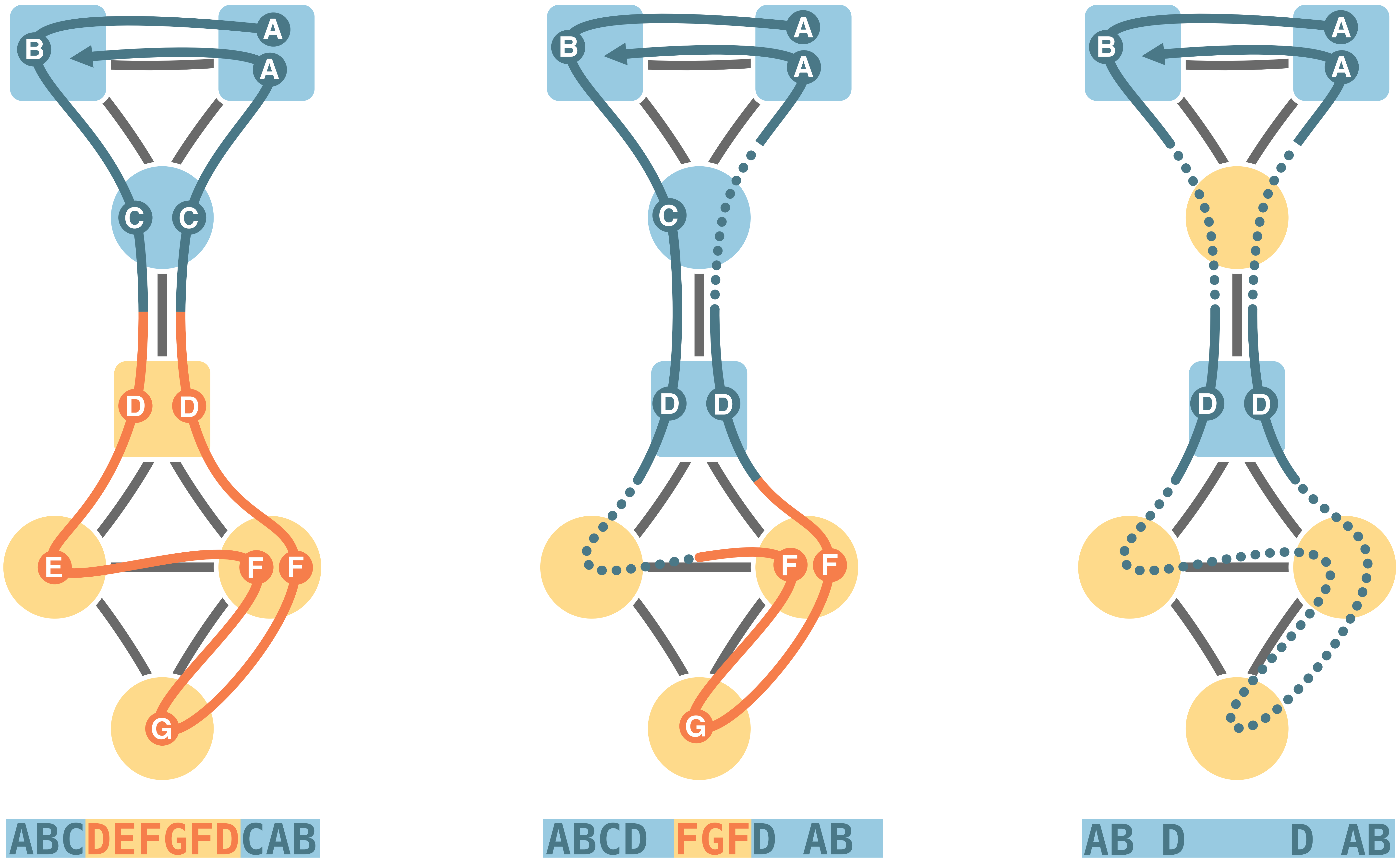
Metadata dependence, $[\mathbf{p}, +\infty]$

- If $c > 1$, the walker will encode more frequently at nodes belonging to the same class of the starting node (assortative encoding).
- For $p < c < 1$, encode will be more probable at nodes belonging to a different class than the one of the starting node (disassortative encoding).
- For $c = 1$, the encoding dynamics no longer depend on class assignments (neutral encoding).
- When $p \ll 1$, the structure is irrelevant and absorption depends only on metadata.

Mapping network flows

Random walks with metadata-dependent encoding probabilities

$c \in [1, 3]$ $\quad$ $c \in (3, 10]$ $\quad$ $c > 10$



Mapping network flows

Random walks with metadata-dependent encoding probabilities

Transition matrix

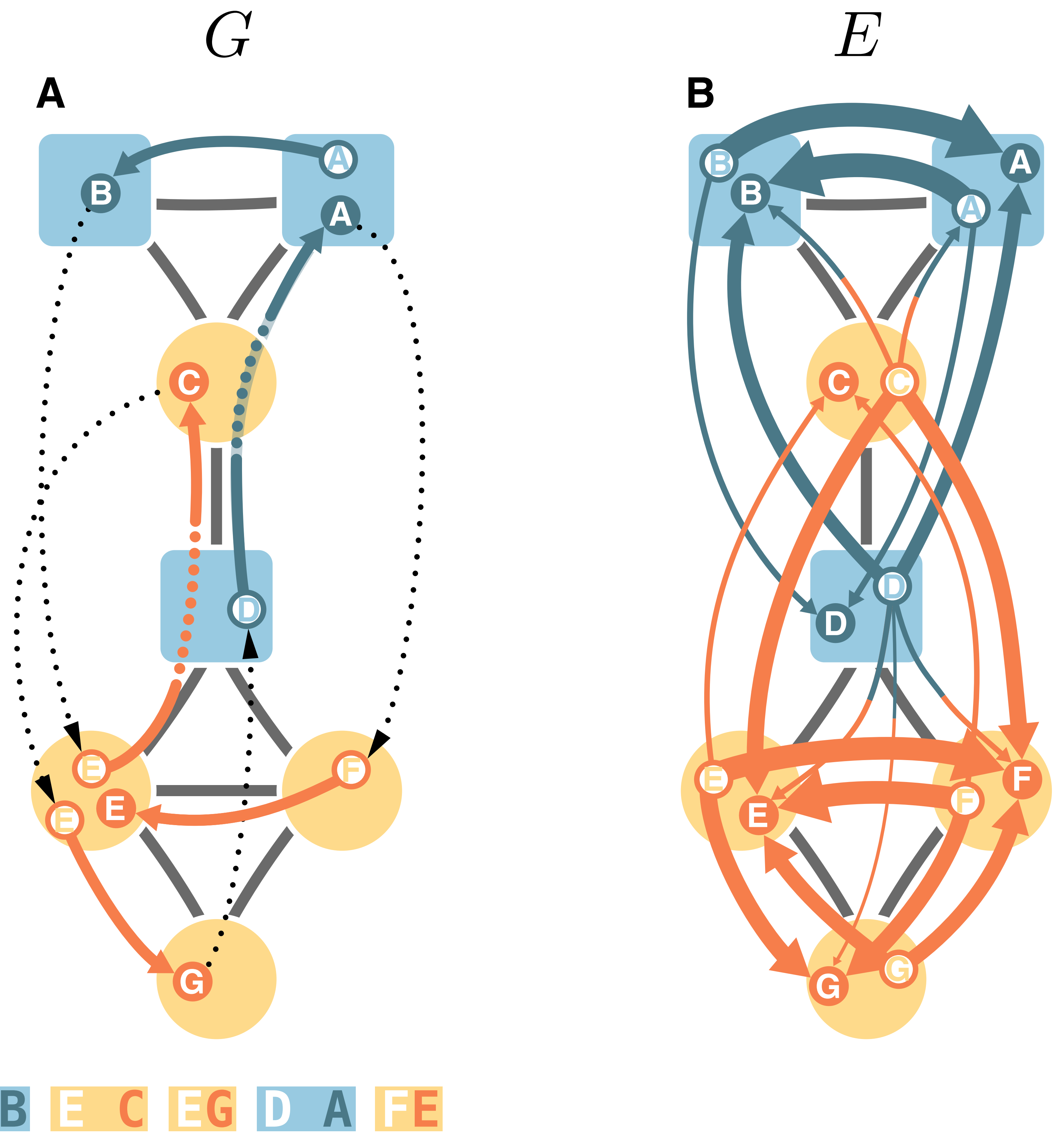
Encoding probabilities

Stationary occupation probability

Original graph

$$G \to E \text{ with } E_i = \sum_{t=1}^{\infty} \mathcal{E}_i^{\intercal} \widetilde{\Pi}_i^t P(0|i),$$

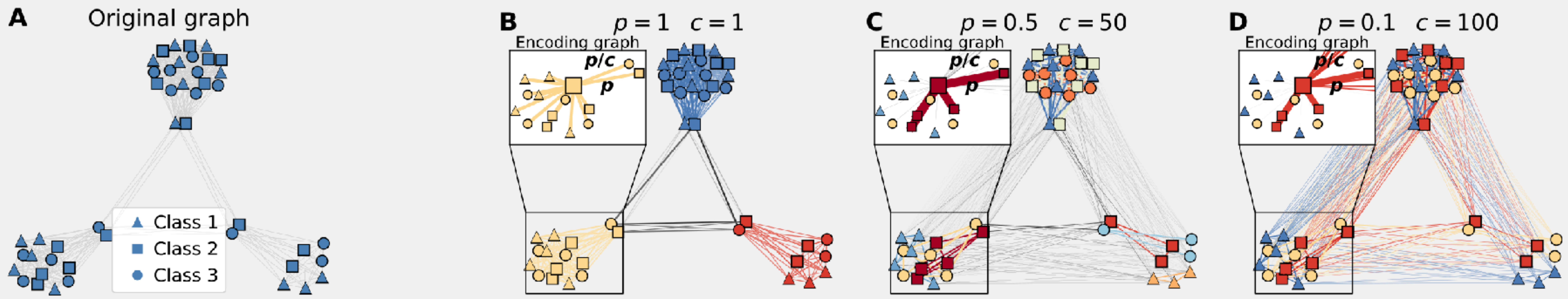
Encoding graph



Mapping network flows

Random walks with metadata-dependent encoding probabilities

Synthetic example



Mapping network flows

Encoding probabilities for real-valued metadata

$$\varepsilon_{ij} = s \exp\left(\frac{-|f_i - f_j|}{b}\right) p + (1 - s)$$

Encoding probability

Real-valued node metadata

Baseline encoding probability

Strength of metadata information

Scale parameter

2. Mapping network flows

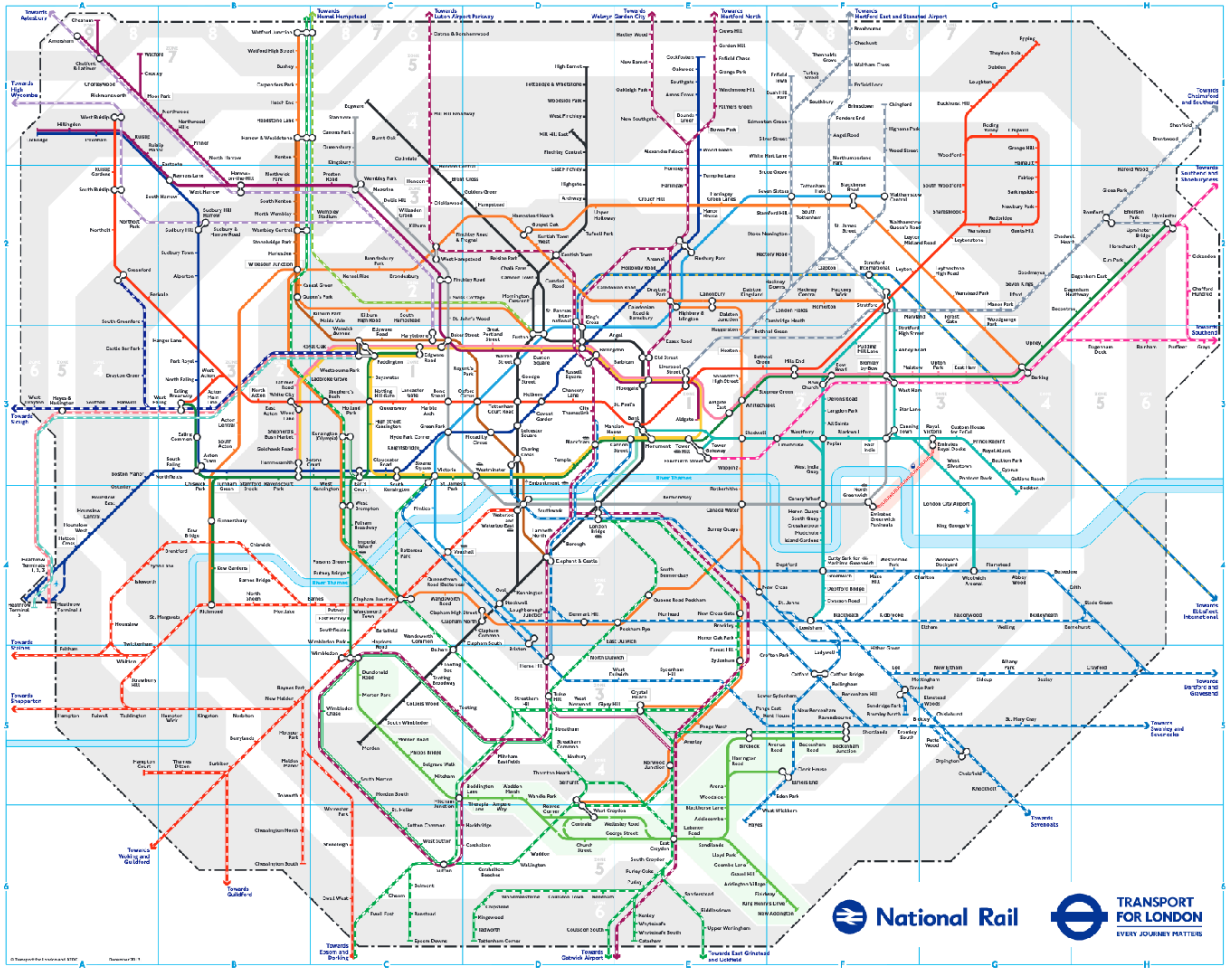
with metadata-dependent encoding

Random walks with metadata-dependent encoding probabilities integrate network information and distant metadata.

3. Mapping network flows

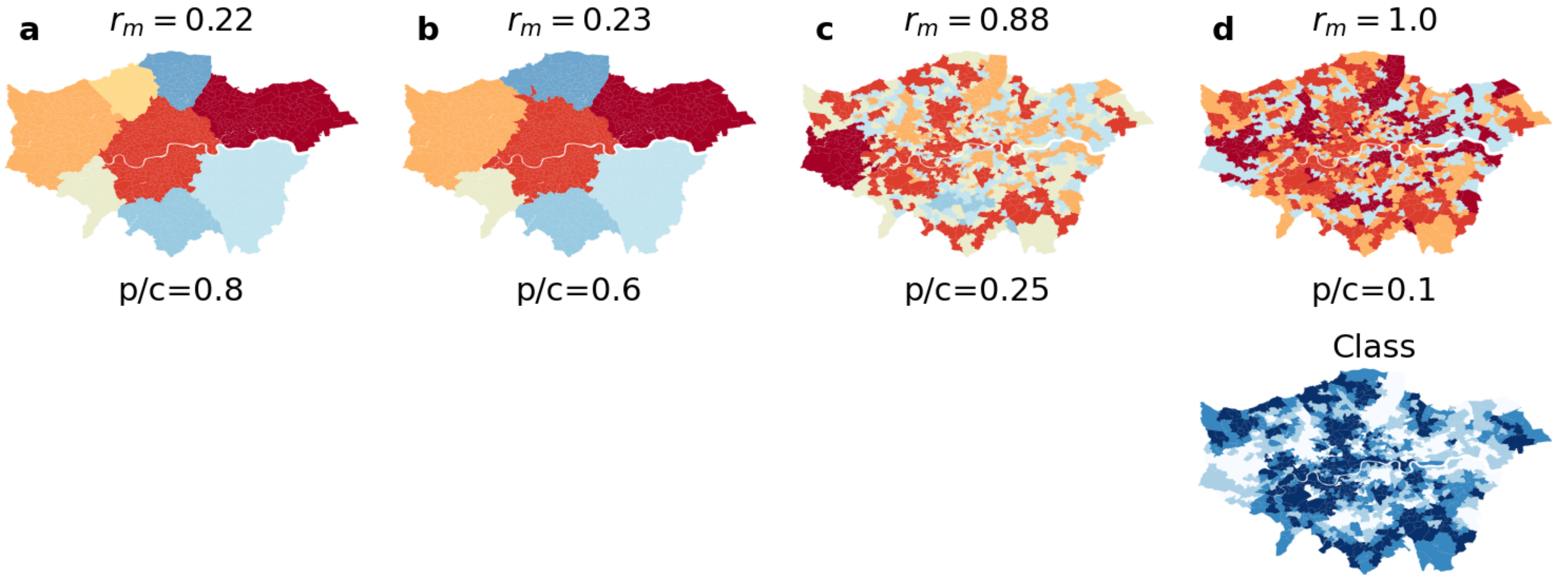
with metadata-dependent encoding

exploits nonlocal relationships

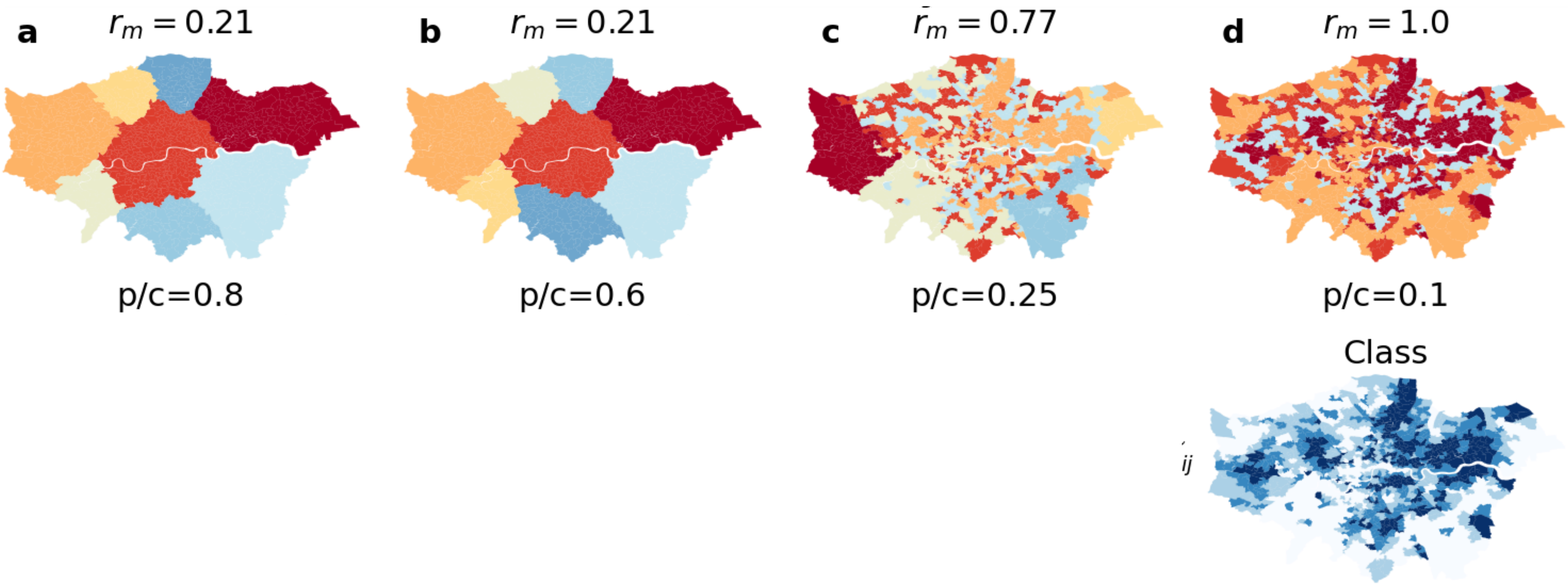

River Thames
National Rail
TRANSPORT FOR LONDON
EVERY JOURNEY MATTERS

Starting point matters

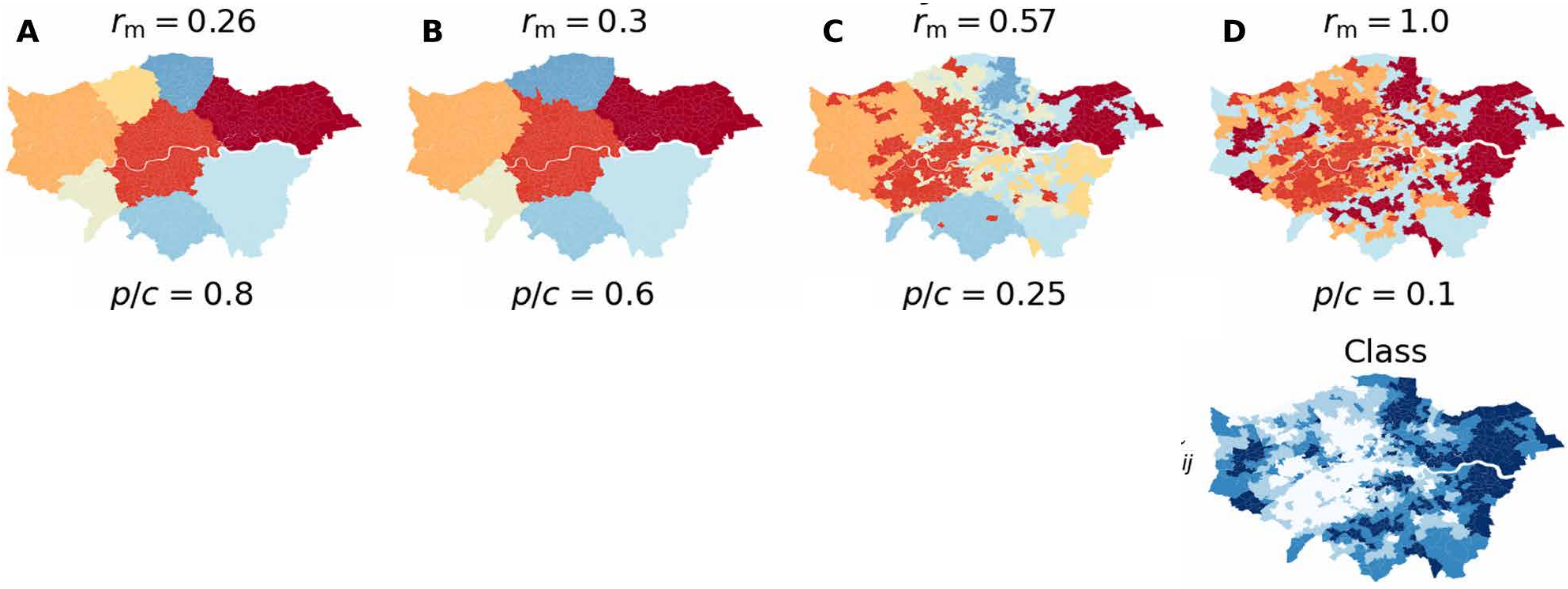
Income class matters in the commuting network of London



Unemployment class matters in the commuting network of London

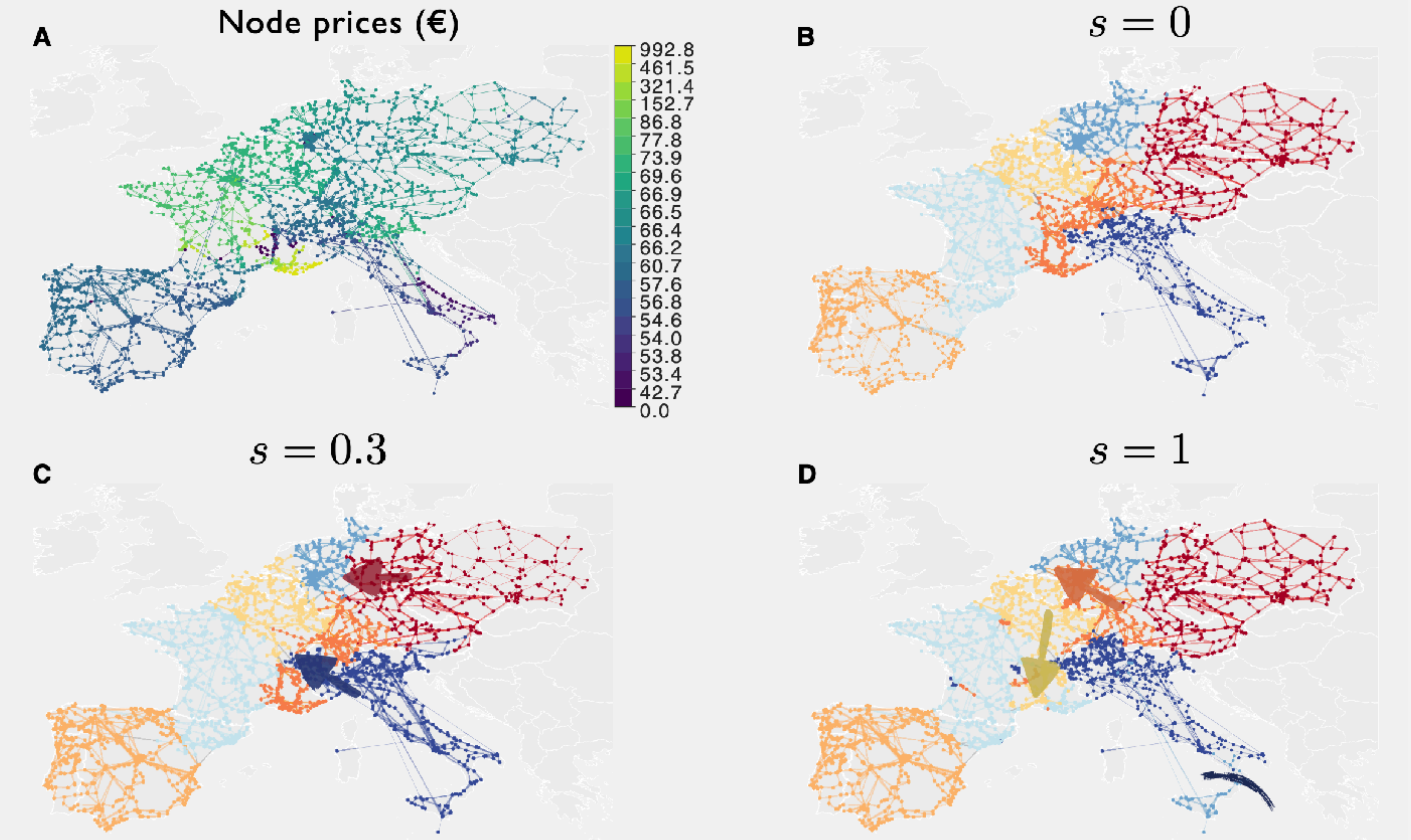


Obesity class matters in the commuting network of London



ENTSO-E Transmission System Map

European power grid network with node prices and optimal partitions



Full derivations and more examples in Sci. Adv. 8, eabn7558 (2022)

SCIENCE ADVANCES | RESEARCH ARTICLE

NETWORK SCIENCE

Mapping nonlocal relationships between metadata and network structure with metadata-dependent encoding of random walks

Aleix Bassolas[1,2,3]†, Anton Holmgren[4]†, Antoine Marot[5], Martin Rosvall[4], Vincenzo Nicosia[1]*

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Integrating structural information and metadata, such as gender, social status, or interests, enriches networks and enables a better understanding of the large-scale structure of complex systems. However, existing approaches to augment networks with metadata for community detection only consider immediately adjacent nodes and cannot exploit the nonlocal relationships between metadata and large-scale network structure present in many spatial and social systems. Here, we develop a flow-based community detection framework based on the map equation that integrates network information and metadata of distant nodes and reveals more complex relationships. We analyze social and spatial networks and find that our methodology can detect functional metadata-informed communities distinct from those derived solely from network information or metadata. For example, in a mobility network of London, we identify communities that reflect the heterogeneity of income distribution, and in a European power grid network, we identify communities that capture relationships between geography and energy prices beyond country borders.

INTRODUCTION

The network structure of a complex system provides meaningful insights into its function, dynamics, and evolution (*1*–*3*). For exam-

and network structure exist, the presence of metadata adds no value to the extended stochastic block models (*16*, *17*). Similarly, encoding metadata in flow-based modules without local correlations

Downloaded fro

3. Mapping network flows

with metadata-dependent encoding

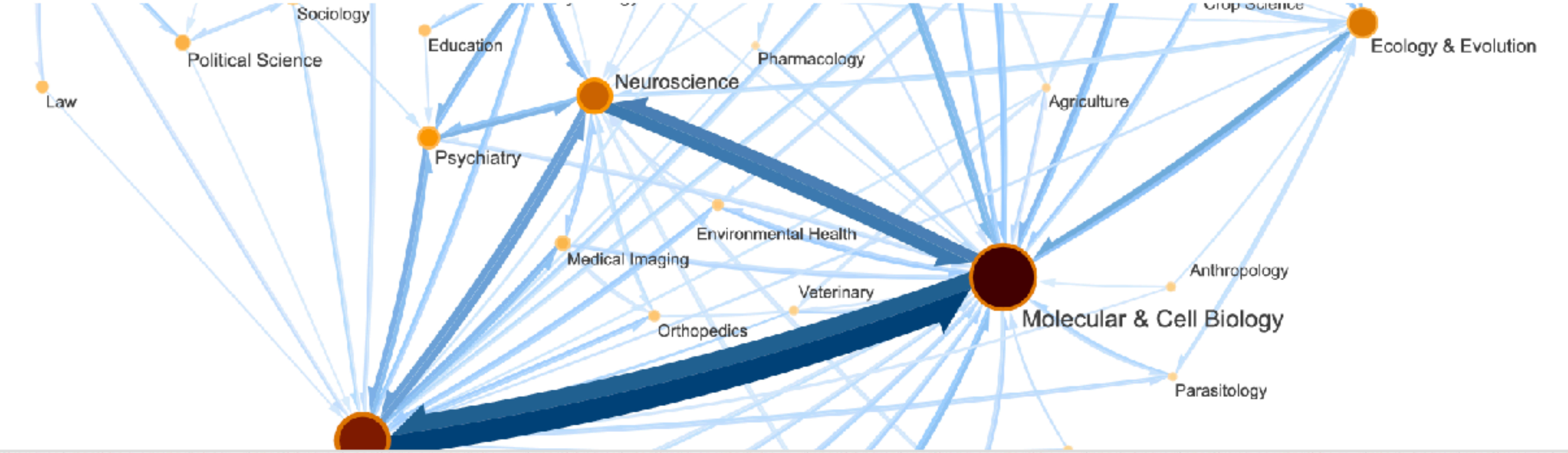
exploits nonlocal relationships

Random walks with metadata-dependent encoding probabilities reveals functional metadata-informed communities

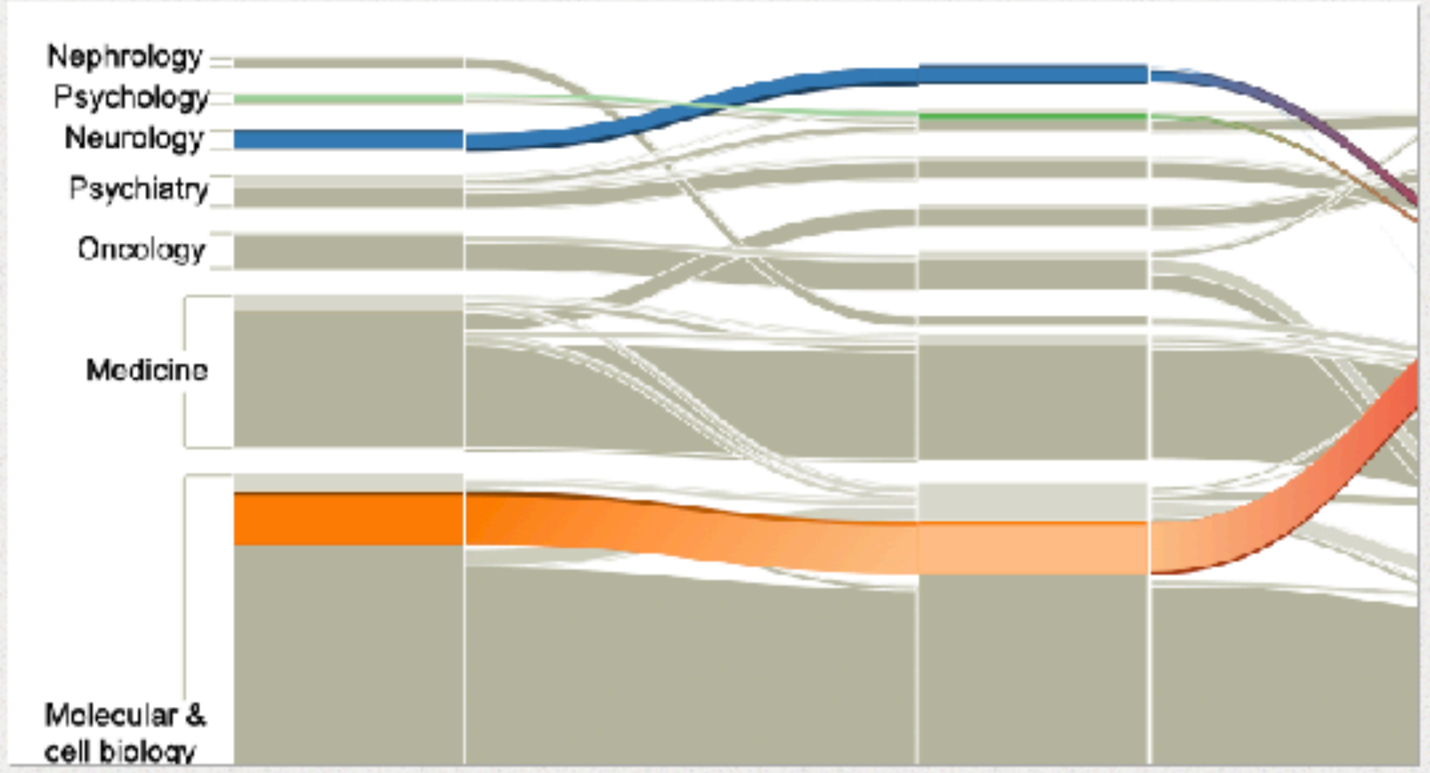
CONCLUSION

Random walks with metadata-dependent encoding probabilities integrate structural and metadata information beyond nodes' immediate neighbors, revealing functional metadata-informed communities

Simplify and highlight important structures in complex networks



Apps »



Code »

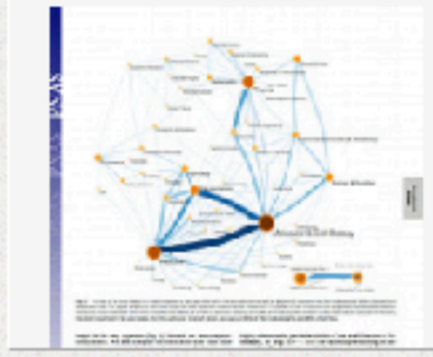
```
from infomap import Infomap
im = Infomap()
im.read_file("ninetriangles.net")
im.add_link(1, 10)
im.run("--two-level --num-trials 5")
print(im.codelength)
for node in im.tree:
  if node.is_leaf:
    print(node.node_id, node.module_id)
```

Publications »

Maps of information flow reveal community structure in complex networks

Martin Rosvall and Carl T. Bergstrom

PNAS **105**, 1118 (2008). [arXiv:0707.0609]



To comprehend the multipartite organization of large-scale biological and social systems, we introduce a new information-theoretic approach to reveal community structure in

News

Jan 29, 2022 Preprint — Map Equation Centrality: A Community-Aware Centrality Score Based on the Map Equation — arXiv:2201.12590

Dec 10, 2021 Research paper — Mapping flows on weighted and directed networks with incomplete observations — J. Comp. Net. 9, 6 (2021)

Three ways to run Infomap

https://www.mapequation.org/infomap/

Python (C++ speed)

To install, run

```
pip install infomap
```

To upgrade, run

```
pip install --upgrade infomap
```

Infomap only supports Python 3

We currently build packages for Python 3.6 to 3.10.

C++

```
git clone git@github.com:mapequation/infomap.git
cd infomap
make -j
```

	OPENMP	WITHOUT OPENMP
WINDOWS	infomap-win.zip	infomap-win-noomp.zip
UBUNTU 18.04	infomap-ubuntu.zip	infomap-ubuntu-noomp.zip
MACOS 10.15	infomap-mac.zip	infomap-mac-noomp.zip

Infomap online (JS speed)

Load network
Edit network or load file

Run Inf
Toggle parameters

Load network

--clu --ftree

```
#source target [weight]
1 2
1 3
1 4
2 1
2 3
3 2
3 1
4 1
4 5
4 6
5 4
5 6
6 5
6 4
```

Load network by dragging & dropping.
Supported formats.

Infomap output will be printed here