Realistic clustering patterns in directed geometric networks

Antoine Allard

Université Laval, Québec, Canada

antoineallard.info

@all_are

Sentinelle Nord

Université Laval

Canada First Research Excellence Fund | Apogée Canada Fonds d'excellence en recherche

NSERC CRSNG

Fonds de recherche Nature et technologies Québec

compute canada | calcul canada

Calcul Québec

Network models

Why?

- Mathematical representation -> analytical results and predictions.
- Identify the mechanisms behind a set of topological properties.
- Disentangle the effect of various topological properties (e.g. assortative mixing vs. clustering on the percolation threshold [1]).
- Identify significant patterns of connection in real networks (i.e. null models).
- Perform *in silico* controlled experiments (e.g. simulation of epidemic spreading).
- …

[1] Phys. Rev. E 80, 020901 (2009)
[2] SIAM Rev. 60, 315 (2018)
[3] Phys. Rev. Lett. 89, 208701 (2002)
[4] Phys. Rev. X 9, 011023 (2019)
[5] Soc. Networks 5, 109 (1983)
[6] Appl. Netw. Sci. 4, 122 (2019)
[7] Nature 393, 440 (1998)
[8] SIAM Rev. 45, 167 (2003)

Network models

Why?

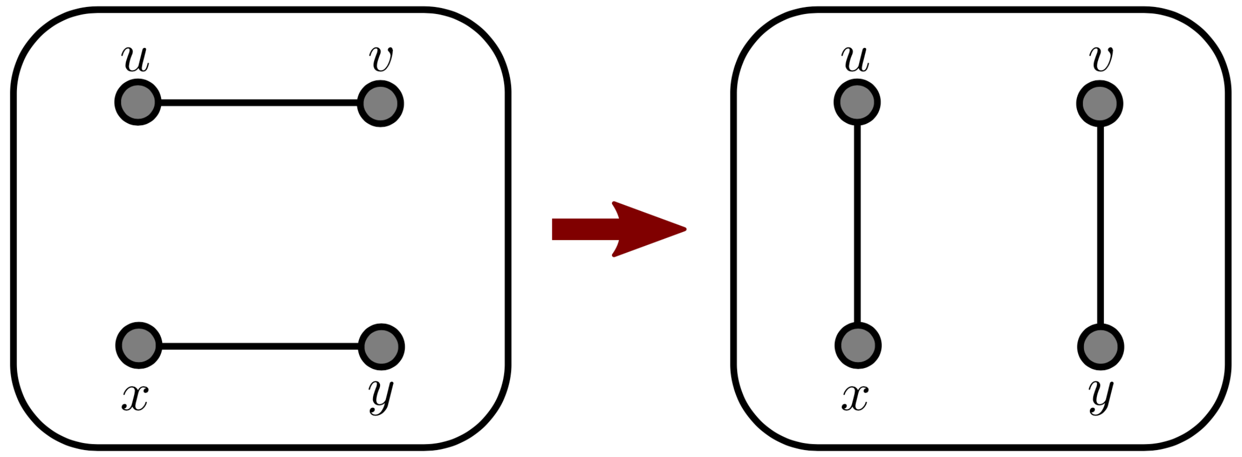
- Mathematical representation -> analytical results and predictions.
- Identify the mechanisms behind a set of topological properties.
- Disentangle the effect of various topological properties (e.g. assortative mixing vs. clustering on the percolation threshold [1]).
- Identify significant patterns of connection in real networks (i.e. null models).
- Perform *in silico* controlled experiments (e.g. simulation of epidemic spreading).
- …

Some examples of *equilibrium* (fixed size) network models

Configuration model (and variations)

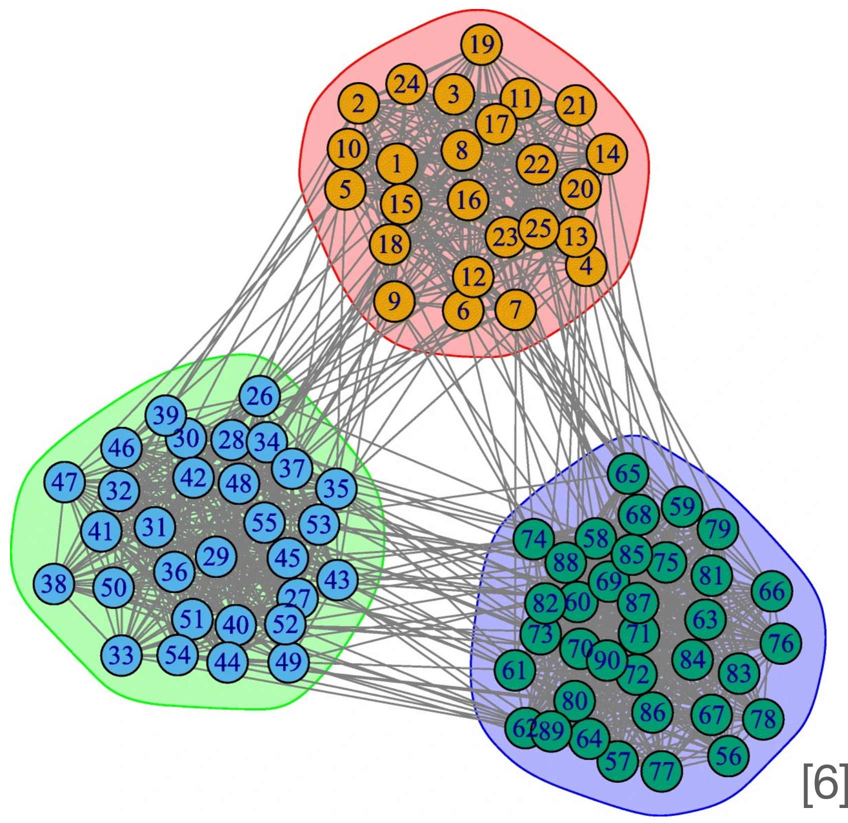
- degree sequence/distribution [2]
- degree-degree correlations [3]
- k-core/onion decomposition [4]

$(u, v), (x, y) \rightsquigarrow (u, x), (v, y)$ [2]



Stochastic block models

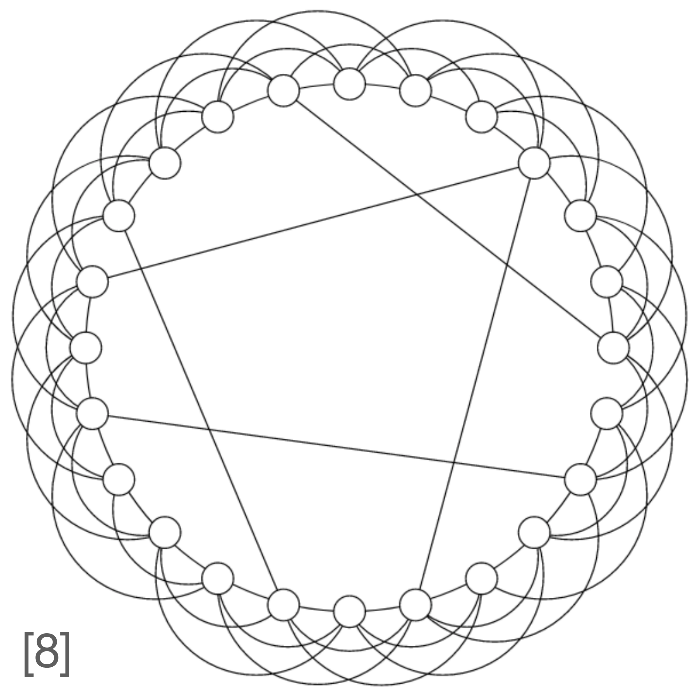
- community structure/detection [5]



[6]

Watts-Strogatz model

- small-world effect [7]



[8]

[1] Phys. Rev. E 80, 020901 (2009)
[2] SIAM Rev. 60, 315 (2018)
[3] Phys. Rev. Lett. 89, 208701 (2002)
[4] Phys. Rev. X 9, 011023 (2019)
[5] Soc. Networks 5, 109 (1983)
[6] Appl. Netw. Sci. 4, 122 (2019)
[7] Nature 393, 440 (1998)
[8] SIAM Rev. 45, 167 (2003)

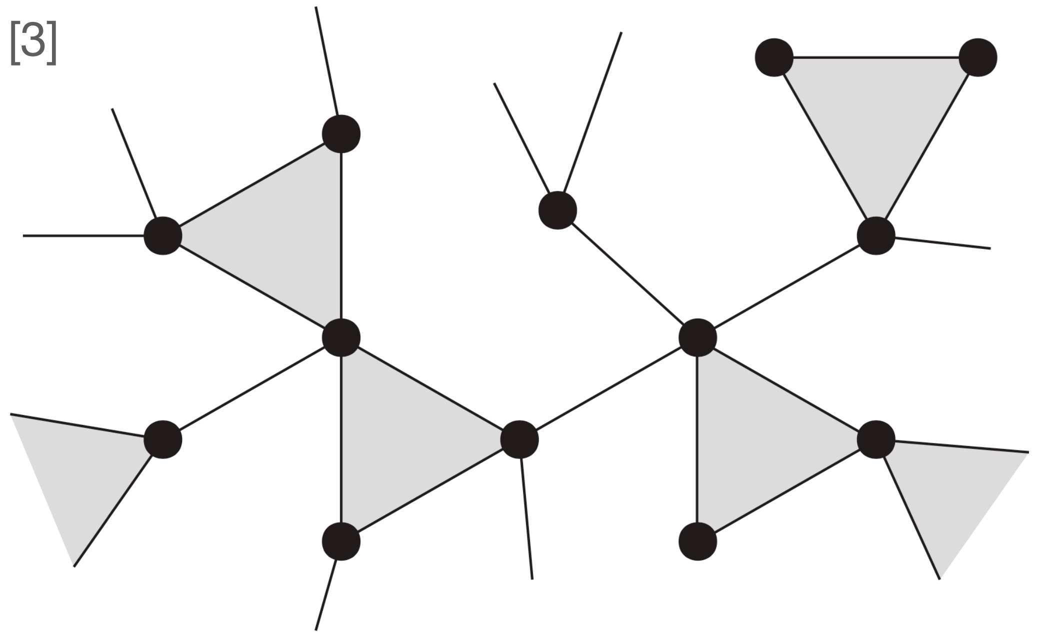
Modeling clustering

Trickier because clustering consists in three-node interactions while our mathematical tools rely on pairwise interactions either explicitly or implicitly.

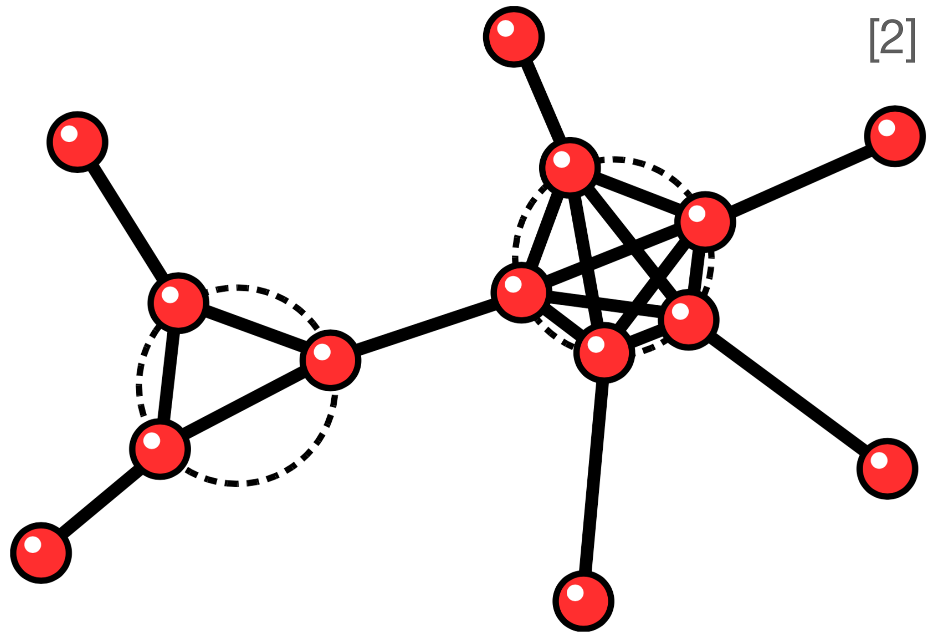
Most models therefore assume

- an underlying tree-like structure
- the networks to be dense

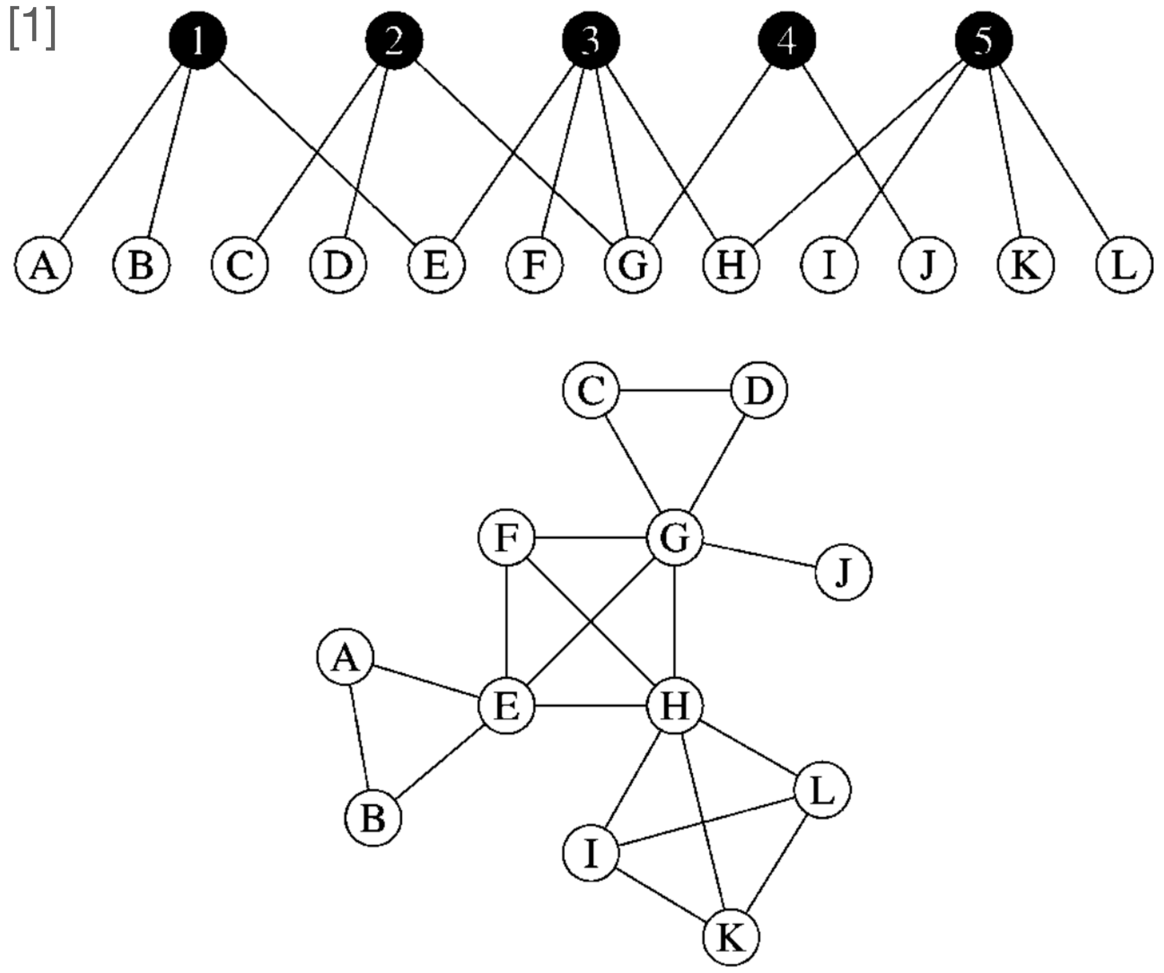
[3]



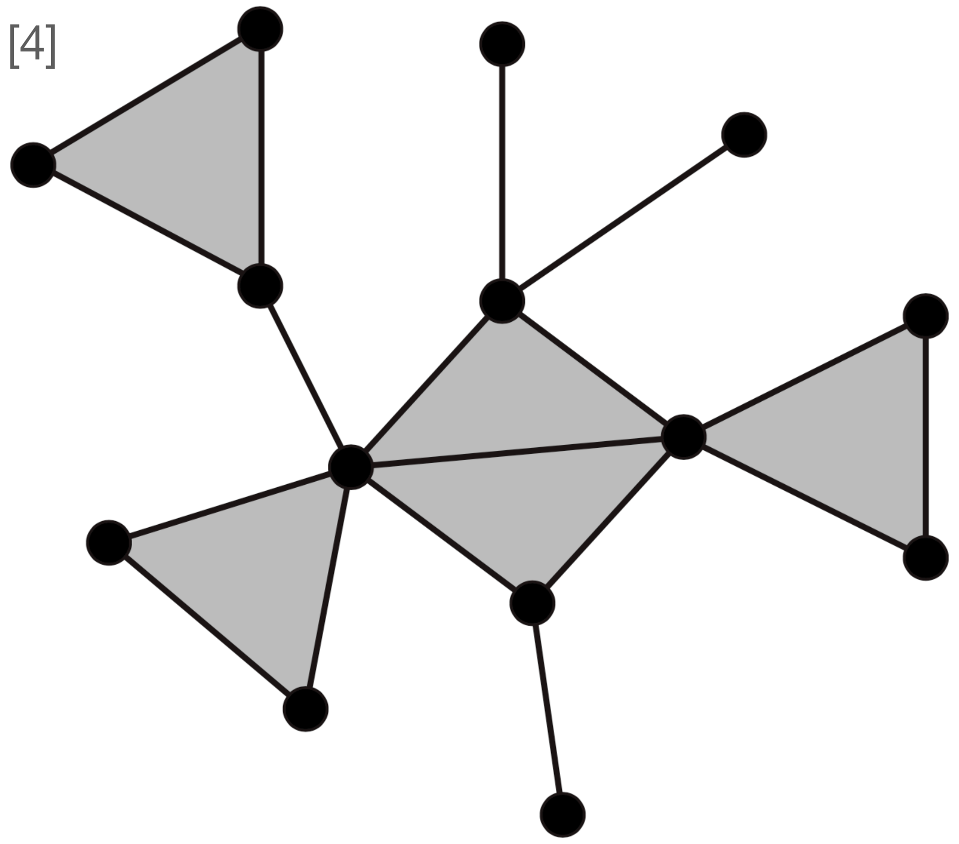
[2]

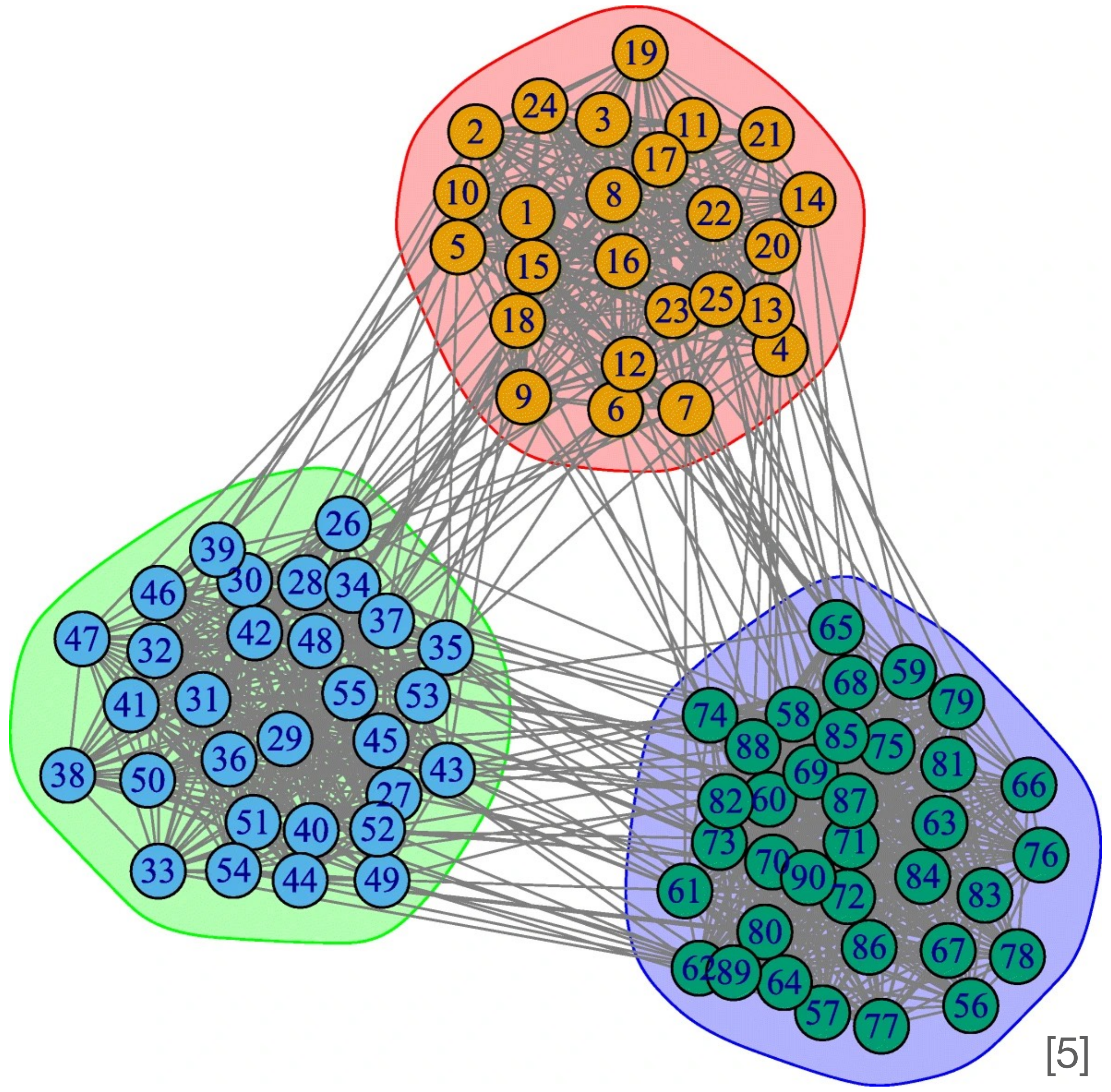


[1]



[4]

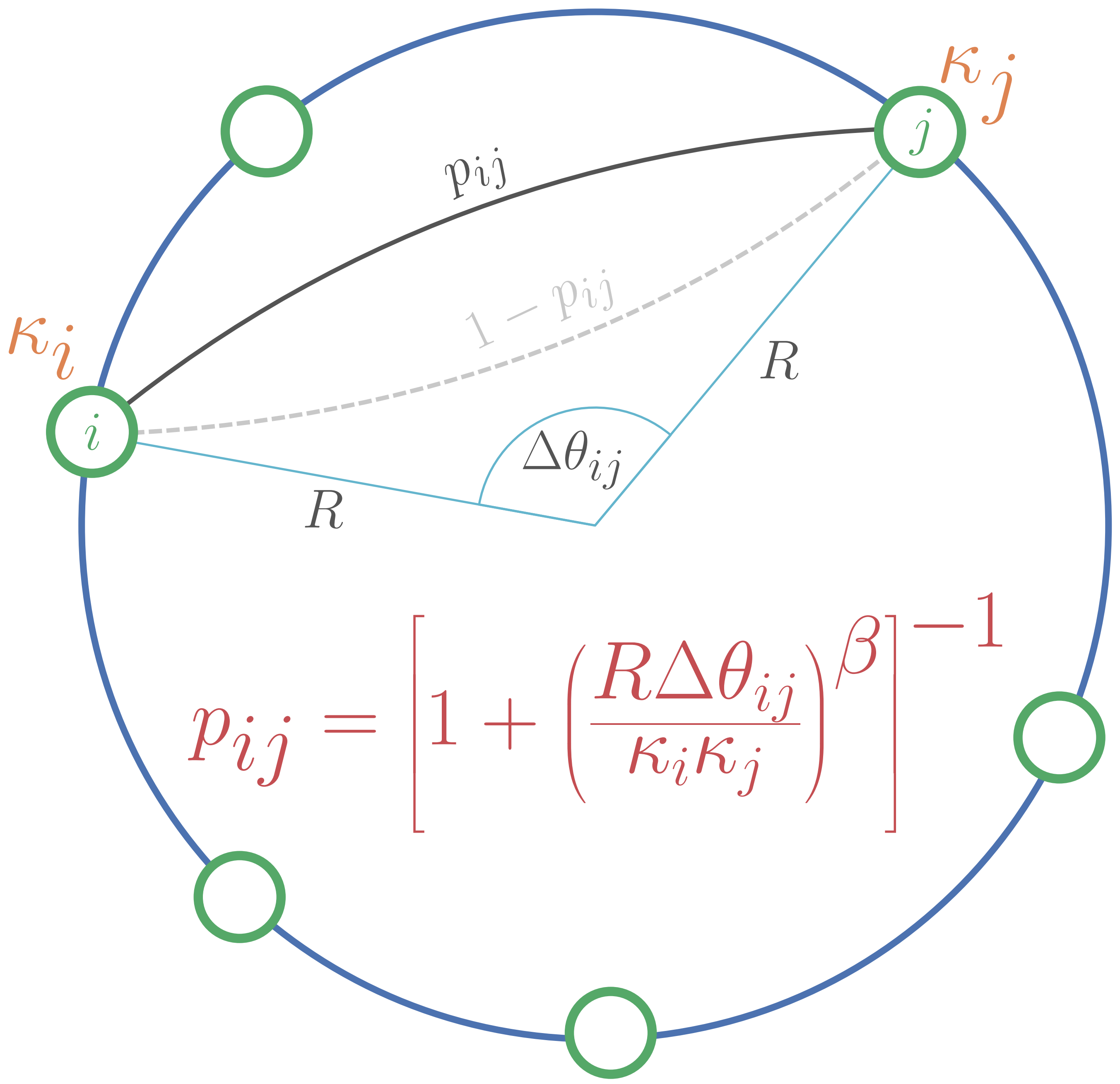




[5]

[1] Phys. Rev. E 68, 026121 (2003)
[2] Phys. Rev. E 80, 036107 (2009)
[3] Phys. Rev. Lett. 103, 058701 (2009)
[4] Phys. Rev. E 82, 066118 (2010)
[5] Appl. Netw. Sci. 4, 122 (2019)

A geometric approach to clustering: the $\mathbb{S}^1/\mathbb{H}^2$ model



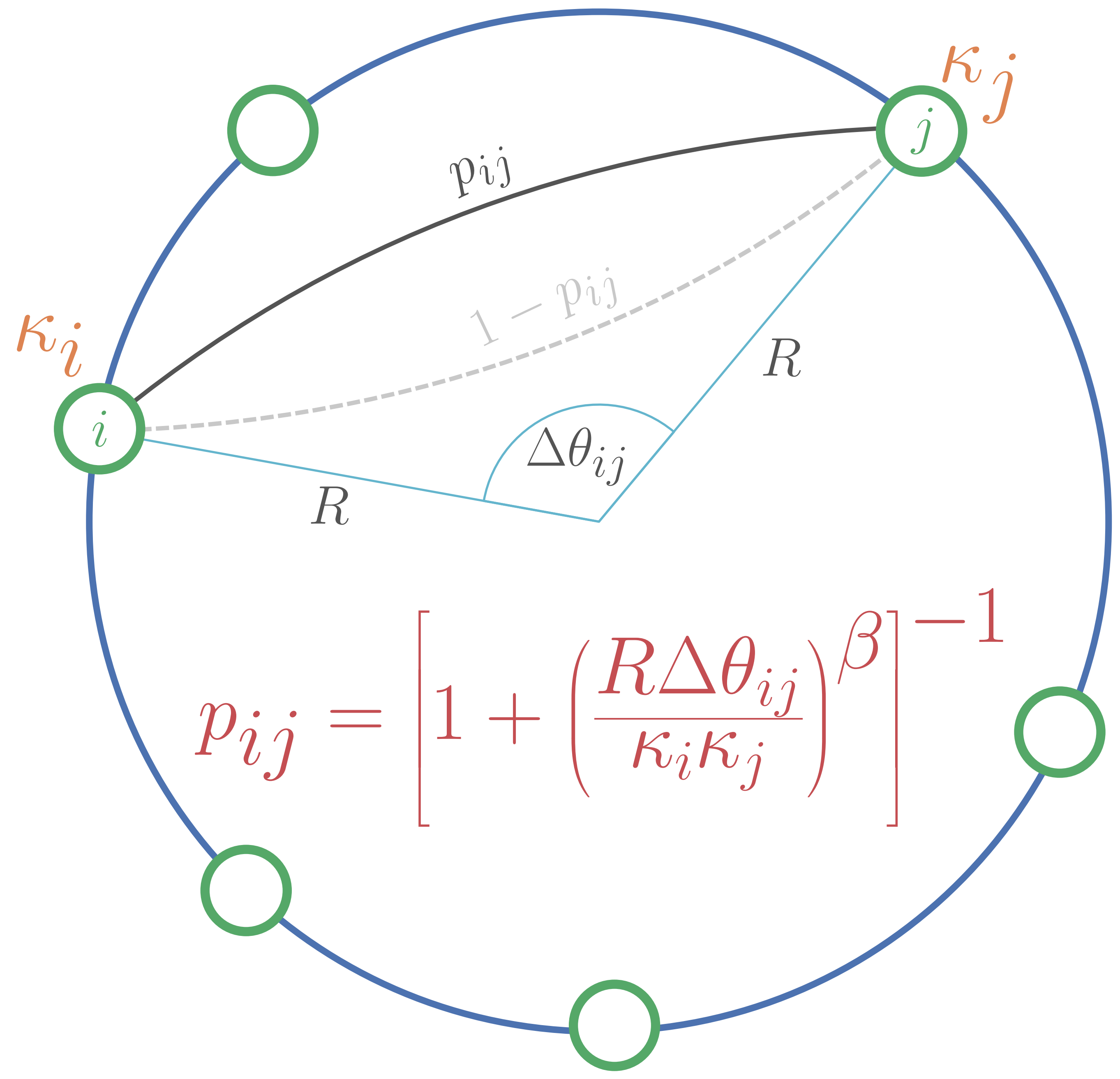
The $\mathbb{S}^1$ model

1. Sprinkle N nodes uniformly on a circle of radius R.
2. Assign an expected degree κ to each node according to some pdf $\rho(\kappa)$.
3. Draw a link between node i and node j with probability p_{ij}.

★ fixes the expected degree of nodes (κ) ~> soft configuration model (CM)
★ triangle inequality of the underlying metric space ~> triangles from pairwise interactions
★ level of clustering tuned with parameter β

[1] Phys. Rev. E 80, 035101 (2009)
[2] Phys. Rev. E 82, 036106 (2010)
[3] Phys. Rev. Lett. 100, 078701 (2008)
[4] Nat. Rev. Phys. 3, 114 (2021)
[5] Nat. Commun. 8, 14103 (2017)
[6] Phys. Rev. E 84, 026114 (2011)
[7] Phys. Rev. E 95, 032309 (2017)
[8] Mol. Biosyst. 8, 843 (2012)
[9] Nat. Phys. 12, 1076 (2016)
[10] Phys. Rev. Lett. 118, 218301 (2017)
[11] Nature 489, 537 (2012)
[12] Sci. Rep. 5, 9421 (2015)
[13] J. Stat. Phys. 173, 775 (2018)
[14] New J. Phys. 20, 052002 (2018)
[15] New J. Phys. 21, 123033 (2019)
[16] Nat. Commun. 8, 1615 (2017)
[17] Nat. Commun. 1, 62 (2010)
[18] PNAS 117, 20244 (2020)

A geometric approach to clustering: the $\mathbb{S}^1/\mathbb{H}^2$ model



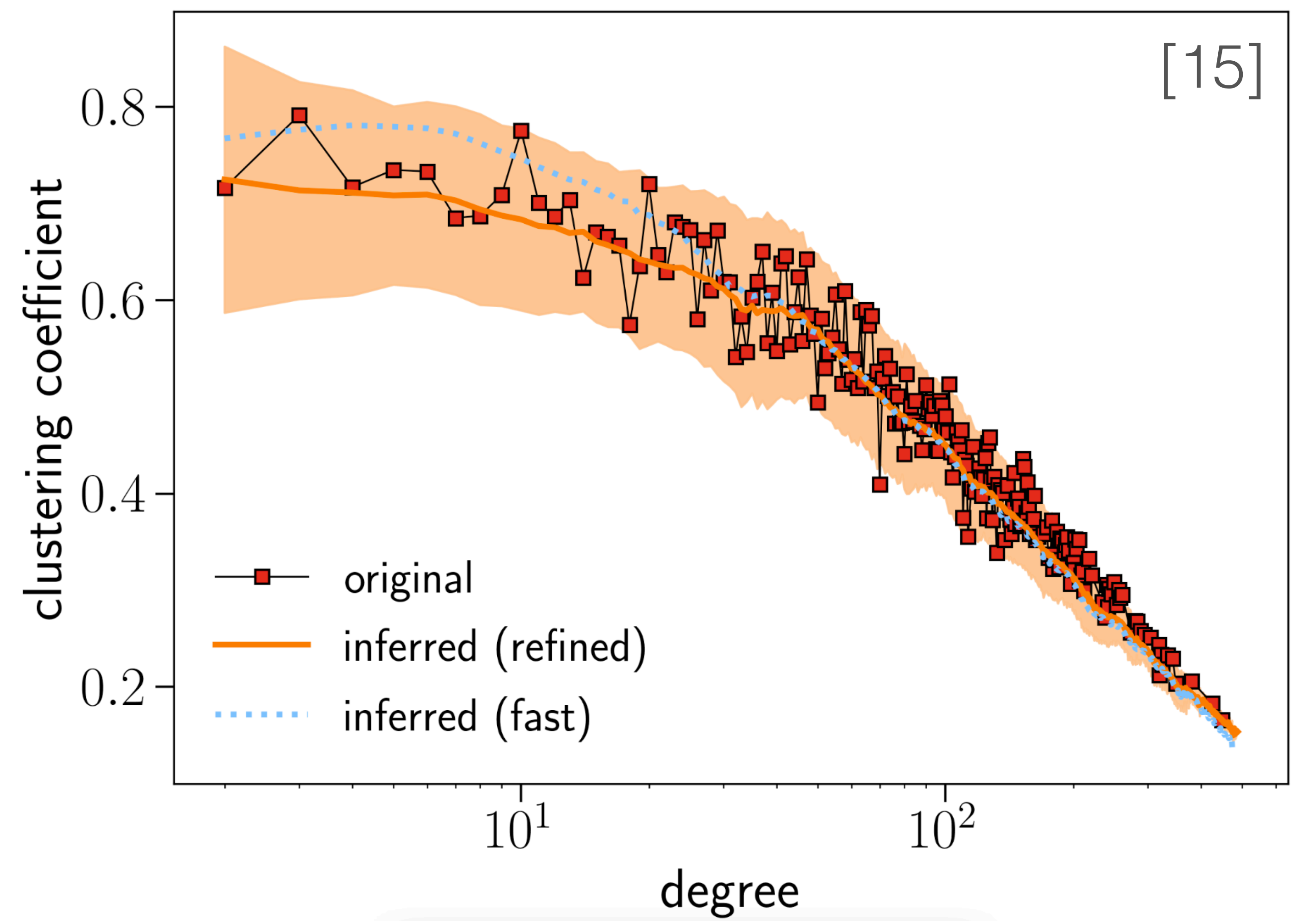
The $\mathbb{S}^1$ model

1. Sprinkle N nodes uniformly on a circle of radius R.
2. Assign an expected degree κ to each node according to some pdf $\rho(\kappa)$.
3. Draw a link between node i and node j with probability p_{ij}.

★ fixes the expected degree of nodes (κ) ~> soft configuration model (CM)

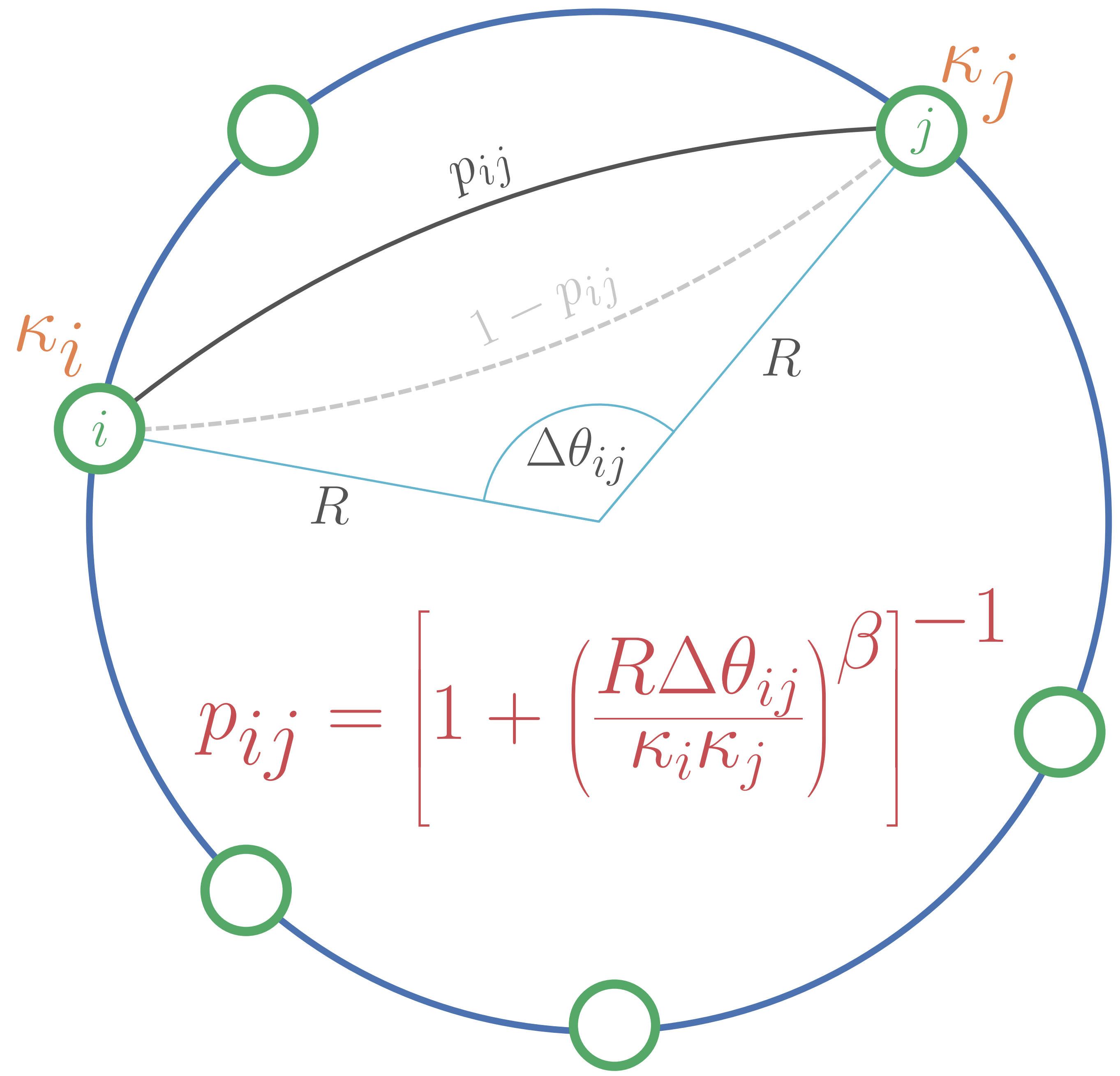
★ triangle inequality of the underlying metric space ~> triangles from pairwise interactions

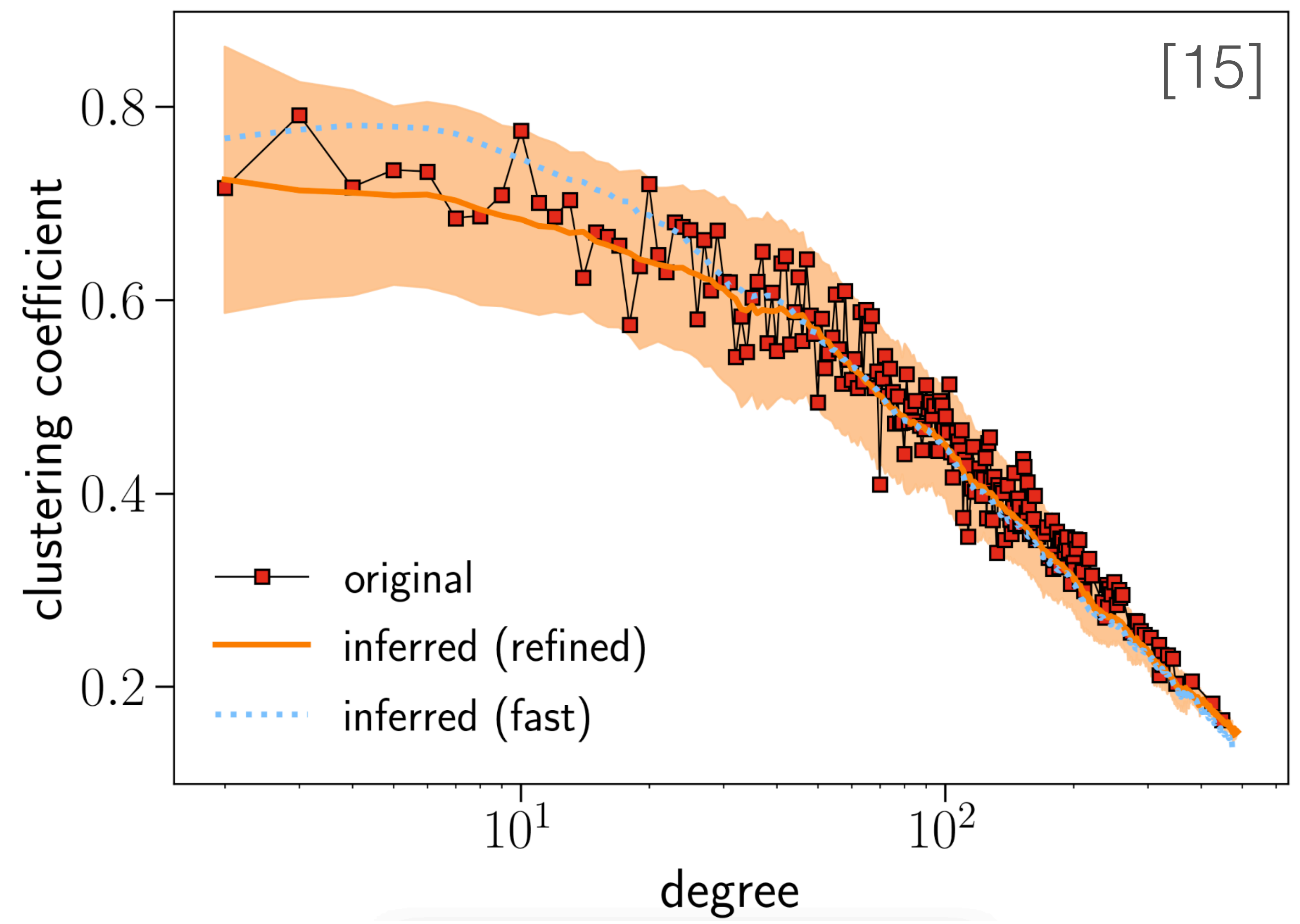
★ level of clustering tuned with parameter β



[1] Phys. Rev. E 80, 035101 (2009)
[2] Phys. Rev. E 82, 036106 (2010)
[3] Phys. Rev. Lett. 100, 078701 (2008)
[4] Nat. Rev. Phys. 3, 114 (2021)
[5] Nat. Commun. 8, 14103 (2017)
[6] Phys. Rev. E 84, 026114 (2011)
[7] Phys. Rev. E 95, 032309 (2017)
[8] Mol. Biosyst. 8, 843 (2012)
[9] Nat. Phys. 12, 1076 (2016)
[10] Phys. Rev. Lett. 118, 218301 (2017)
[11] Nature 489, 537 (2012)
[12] Sci. Rep. 5, 9421 (2015)
[13] J. Stat. Phys. 173, 775 (2018)
[14] New J. Phys. 20, 052002 (2018)
[15] New J. Phys. 21, 123033 (2019)
[16] Nat. Commun. 8, 1615 (2017)
[17] Nat. Commun. 1, 62 (2010)
[18] PNAS 117, 20244 (2020)

A geometric approach to clustering: the $\mathbb{S}^1/\mathbb{H}^2$ model





The $\mathbb{S}^1$ model

1. Sprinkle N nodes uniformly on a circle of radius R.
2. Assign an expected degree κ to each node according to some pdf $\rho(\kappa)$.
3. Draw a link between node i and node j with probability p_{ij}.

★ fixes the expected degree of nodes (κ) ~> soft configuration model (CM)
★ triangle inequality of the underlying metric space ~> triangles from pairwise interactions
★ level of clustering tuned with parameter β

Other properties and generalizations

- Amenable to many analytical calculations
- Geometric interpretation in terms of hyperbolic geometry (the $\mathbb{H}^2$ model) [1,2]
- Parsimonious explanation of self-similarity [3,4]
- Generalizable to weighted [5], bipartite [6,7,8], multiplex [9,10] and growing [11] networks
- Generalizable to networks with community structure [12,13, 14]
- Mapping of real complex networks unto hyperbolic space [15,16]
- Identification of biochemical pathways in E. Coli [8]
- Efficient Internet routing protocols [17]
- Multiscale organization of the human connectome [18]
- Geometrical interpretation of preferential attachment [11]
- …

[1] Phys. Rev. E 80, 035101 (2009)
[2] Phys. Rev. E 82, 036106 (2010)
[3] Phys. Rev. Lett. 100, 078701 (2008)
[4] Nat. Rev. Phys. 3, 114 (2021)
[5] Nat. Commun. 8, 14103 (2017)
[6] Phys. Rev. E 84, 026114 (2011)
[7] Phys. Rev. E 95, 032309 (2017)
[8] Mol. Biosyst. 8, 843 (2012)
[9] Nat. Phys. 12, 1076 (2016)
[10] Phys. Rev. Lett. 118, 218301 (2017)
[11] Nature 489, 537 (2012)
[12] Sci. Rep. 5, 9421 (2015)
[13] J. Stat. Phys. 173, 775 (2018)
[14] New J. Phys. 20, 052002 (2018)
[15] New J. Phys. 21, 123033 (2019)
[16] Nat. Commun. 8, 1615 (2017)
[17] Nat. Commun. 1, 62 (2010)
[18] PNAS 117, 20244 (2020)

Three challenges in modeling directed networks

Properties of any metric space

Identity of indiscernibles	$d(x, y) = 0 \quad \Leftrightarrow \quad x = y$
Non-negativity	$d(x, y) \geq 0$
Symmetry	$d(x, y) = d(y, x)$
Triangle inequality	$d(x, y) \leq d(x, z) + d(z, y)$

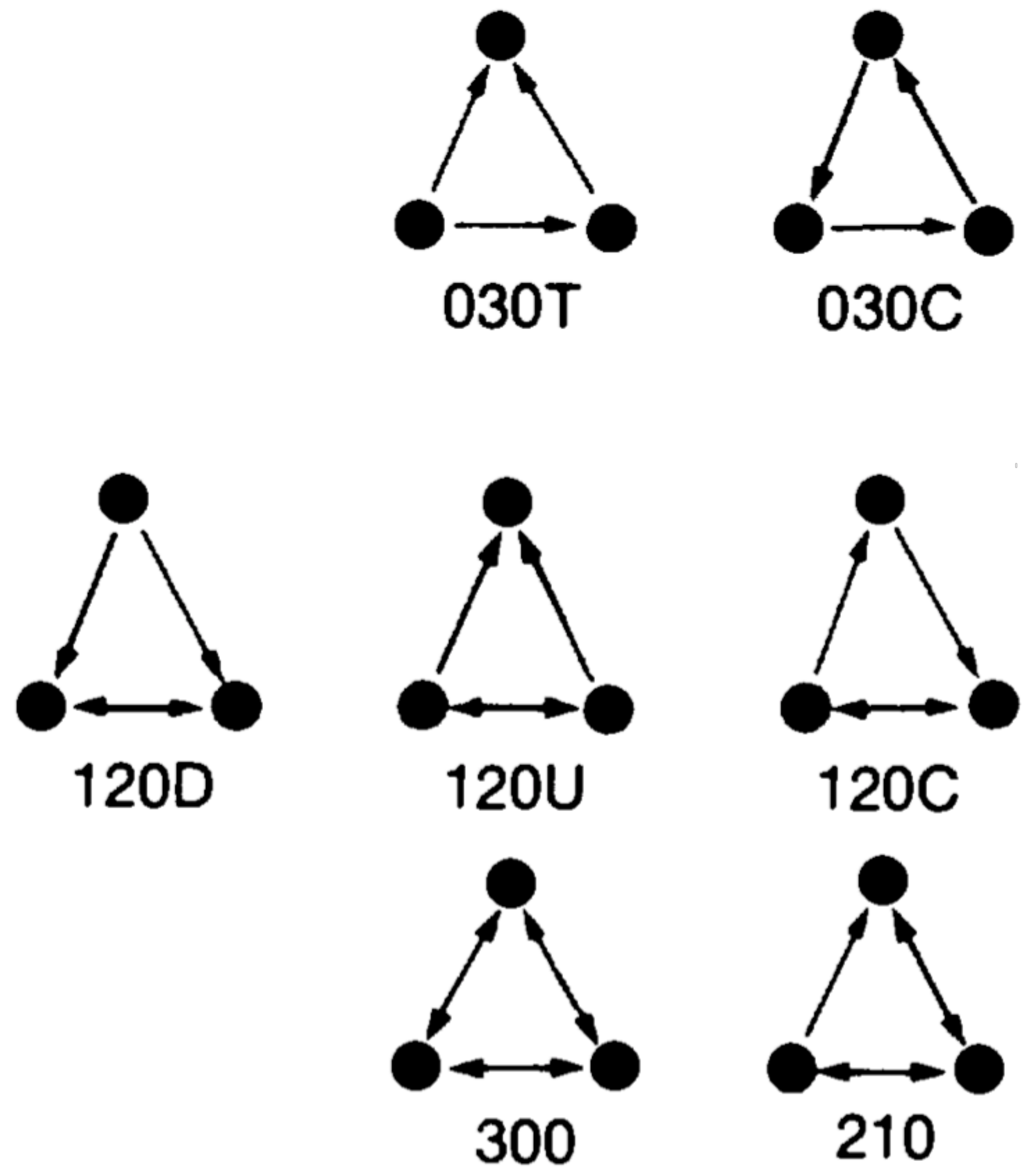
Adapted from Holland & Leinhardt. Local Structure in Social Networks. Sociol. Methodol., 7, 1–45 (1976)

Three challenges in modeling directed networks

Properties of any metric space

Identity of indiscernibles	$d(x, y) = 0 \quad \Leftrightarrow \quad x = y$
Non-negativity	$d(x, y) \geq 0$
Symmetry	$d(x, y) = d(y, x)$
Triangle inequality	$d(x, y) \leq d(x, z) + d(z, y)$

Clustering: 7 cycles of length 3



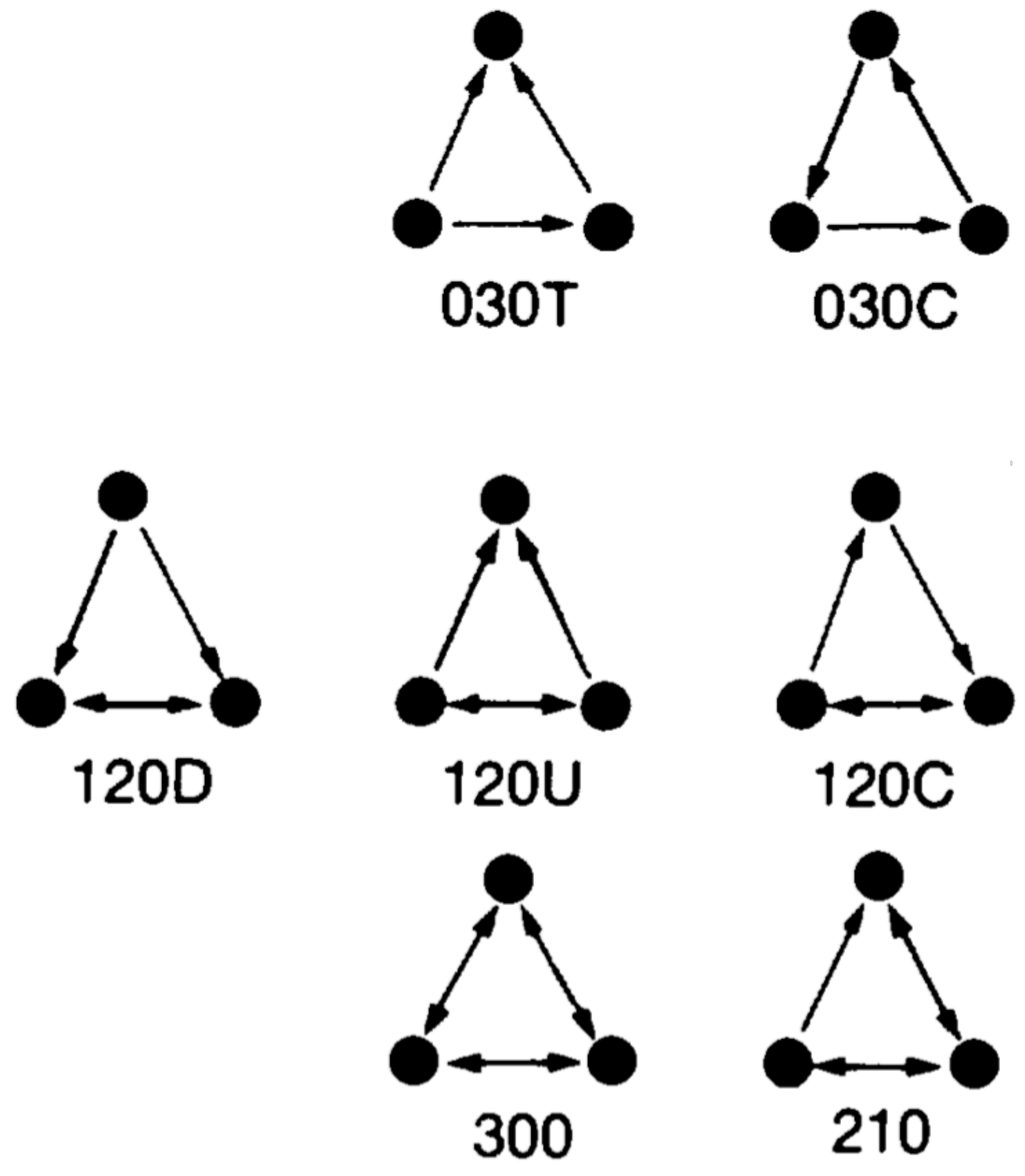
Adapted from Holland & Leinhardt. Local Structure in Social Networks. Sociol. Methodol., 7, 1–45 (1976)

Three challenges in modeling directed networks

Properties of any metric space

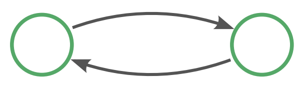
Identity of indiscernibles	$d(x,y) = 0 \quad \Leftrightarrow \quad x = y$
Non-negativity	$d(x,y) \geq 0$
Symmetry	$d(x,y) = d(y,x)$
Triangle inequality	$d(x,y) \leq d(x,z) + d(z,y)$

Clustering: 7 cycles of length 3

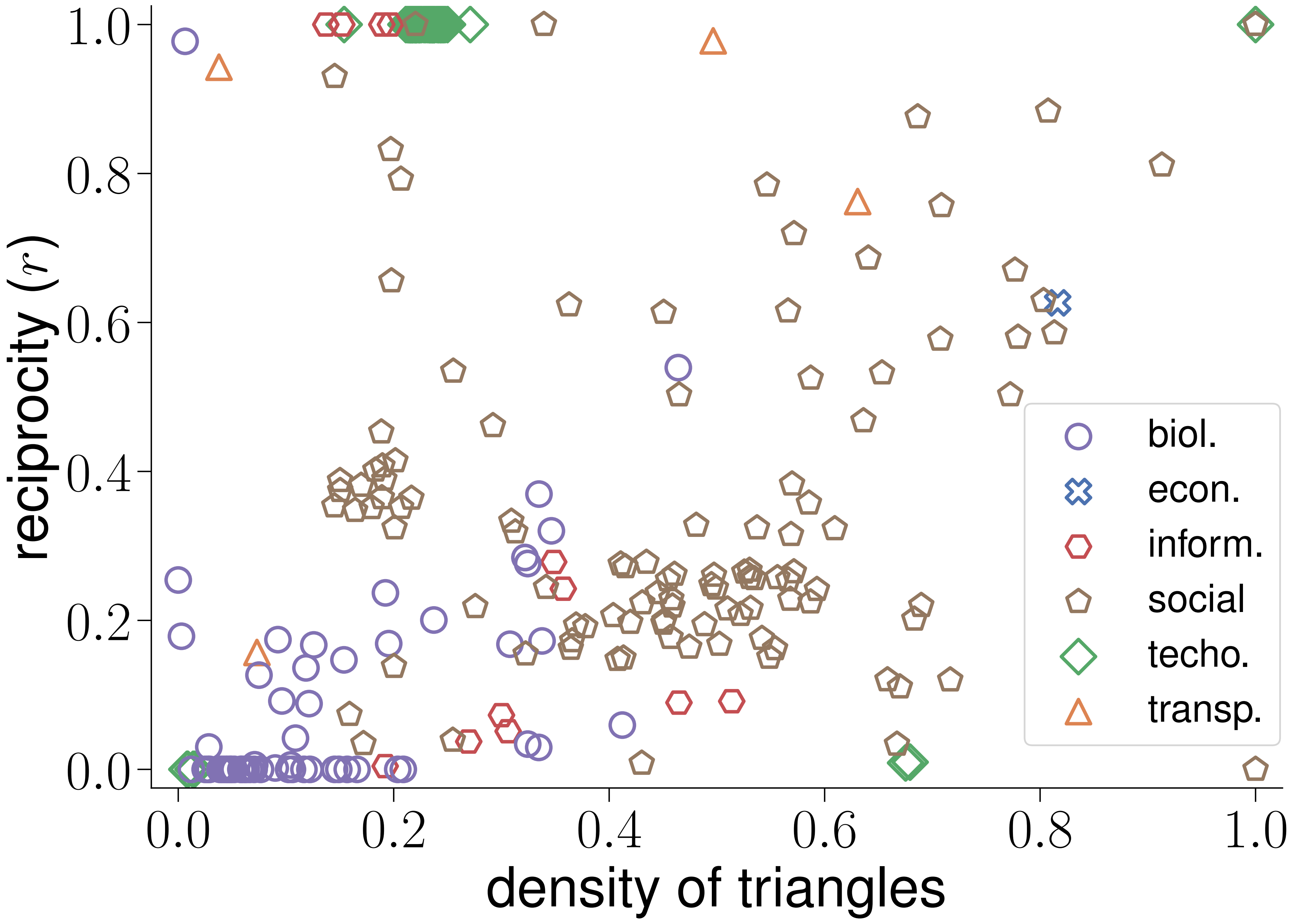


Adapted from Holland & Leinhardt. Local Structure in Social Networks. Sociol. Methodol., 7, 1–45 (1976)

Reciprocity: cycles of length 2

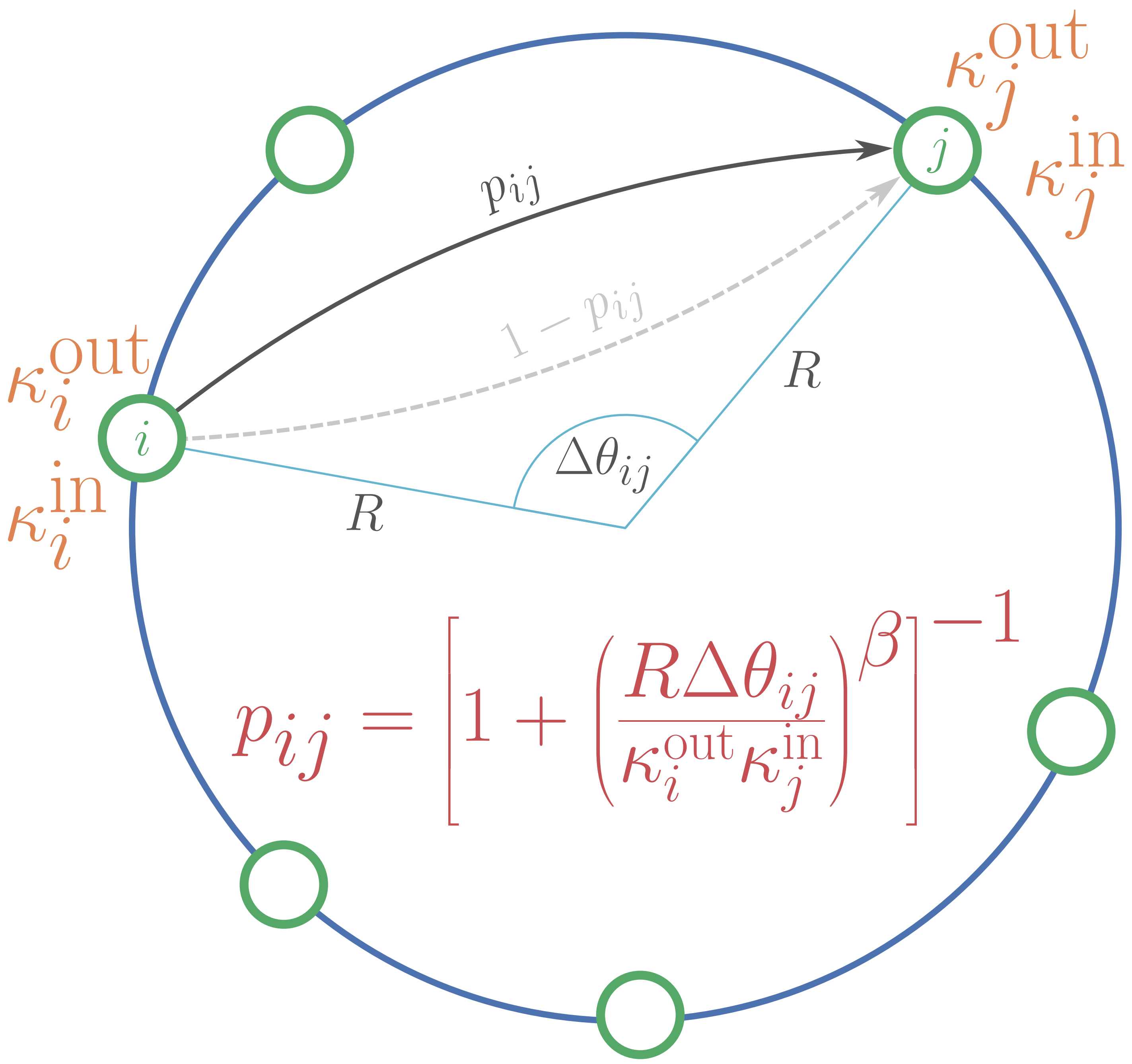


$$r = \frac{L^{\leftrightarrow}}{L} = \frac{[\text{number of reciprocal links}]}{[\text{number of links}]}$$



287 network datasets downloaded from Netzschleuder (networks.skewed.de).

The directed $\mathbb{S}^1$ model



$$p_{ij} = \left[1 + \left(\frac{R\Delta\theta_{ij}}{\kappa_i^{\text{out}}\kappa_j^{\text{in}}}\right)^{\beta}\right]^{-1}$$

The directed $\mathbb{S}^1$ model

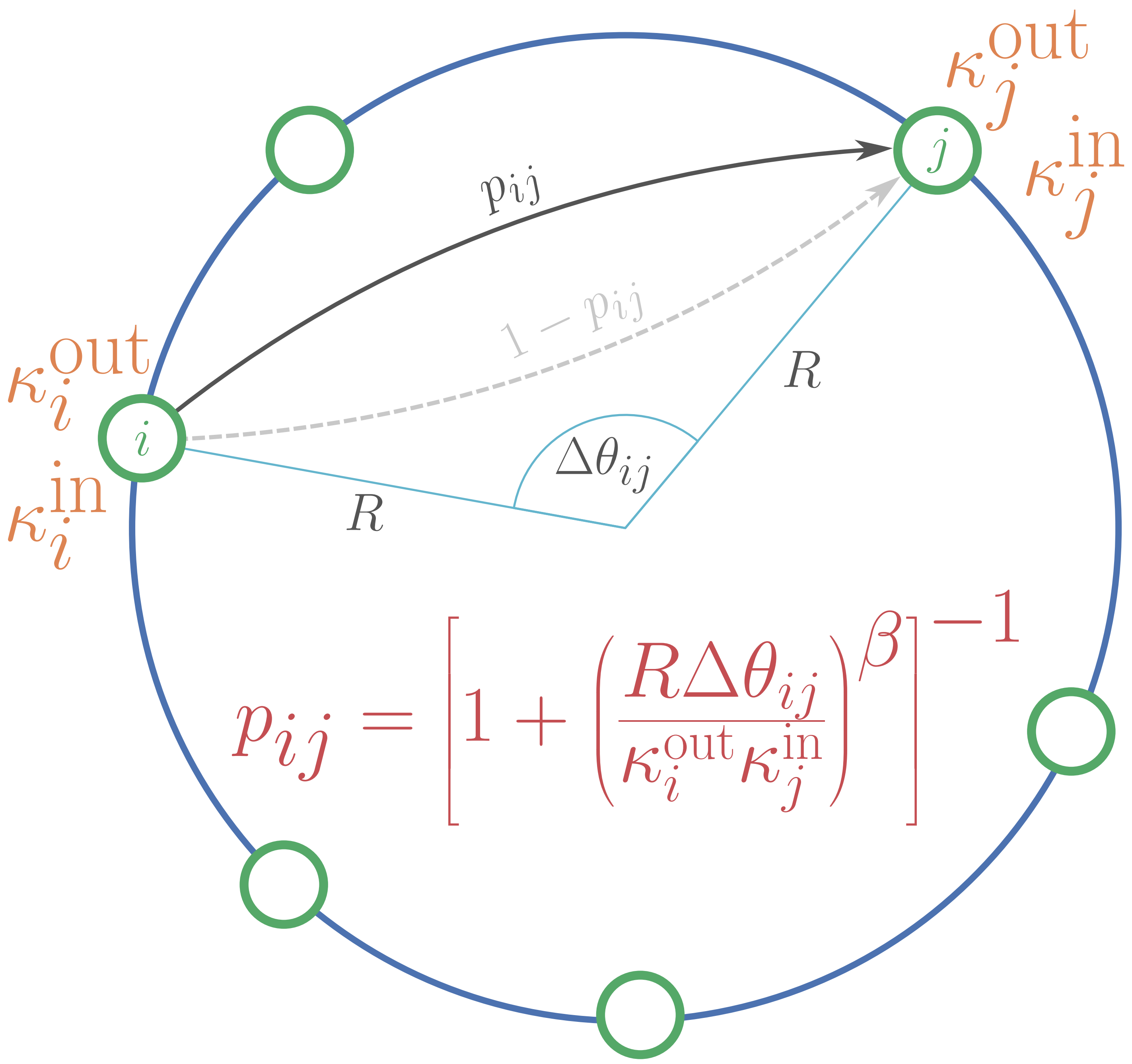
1. Sprinkle N nodes uniformly on a circle of radius R.
2. Assign expected in-degree κ^{in} and out-degree κ^{out} to each node according to some joint pdf $\rho(\kappa^{\text{in}}, \kappa^{\text{out}})$.
3. Draw a link from node i to node j with probability p_{ij}.

★ fixes the expected in-degree and out-degree of nodes $(\kappa^{\text{in}}, \kappa^{\text{out}})$ ~> soft directed CM

★ triangle inequality of the underlying metric space -> triangles from pairwise interactions

★ density of triangles tuned with parameter β

The directed $\mathbb{S}^1$ model



$$p_{ij} = \left[1 + \left(\frac{R\Delta\theta_{ij}}{\kappa_i^{out}\kappa_j^{in}}\right)^{\beta}\right]^{-1}$$

$$\mathbb{E}\left[k^{in}\middle|\kappa^{in}\right] \simeq \kappa^{in}$$

$$\mathbb{E}\left[k^{out}\middle|\kappa^{out}\right] \simeq \kappa^{out}$$

$$P(k^{in}, k^{out}) \simeq \iint \frac{[\kappa^{in}]^{k^{in}} e^{-\kappa^{in}}}{k^{in}!} \frac{[\kappa^{out}]^{k^{out}} e^{-\kappa^{out}}}{k^{out}!} \times \rho(\kappa^{in}, \kappa^{out}) d\kappa^{in} d\kappa^{out}$$

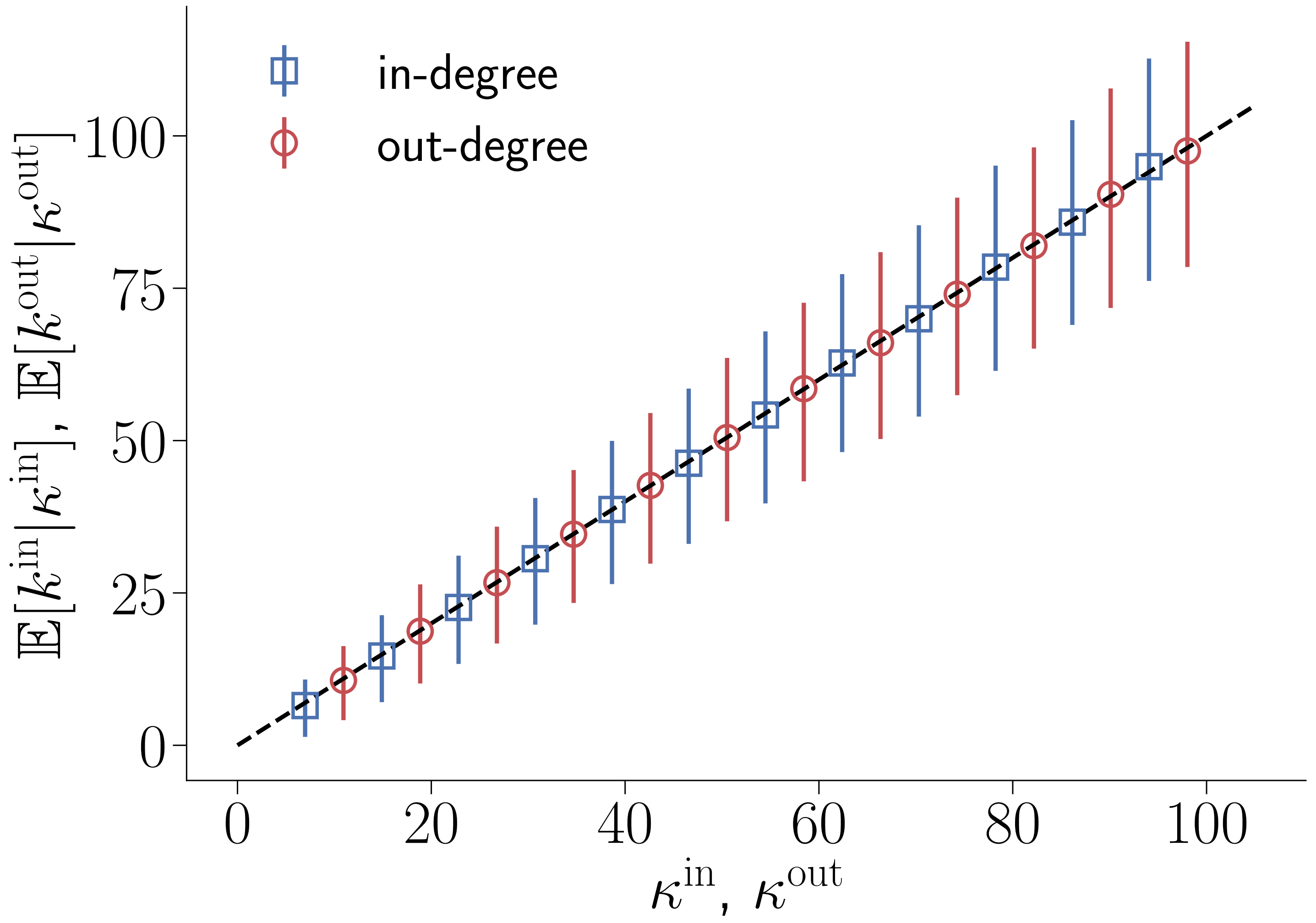
The directed $\mathbb{S}^1$ model

1. Sprinkle N nodes uniformly on a circle of radius R.
2. Assign expected in-degree κ^{in} and out-degree κ^{out} to each node according to some joint pdf $\rho(\kappa^{in}, \kappa^{out})$.
3. Draw a link from node i to node j with probability p_{ij}.

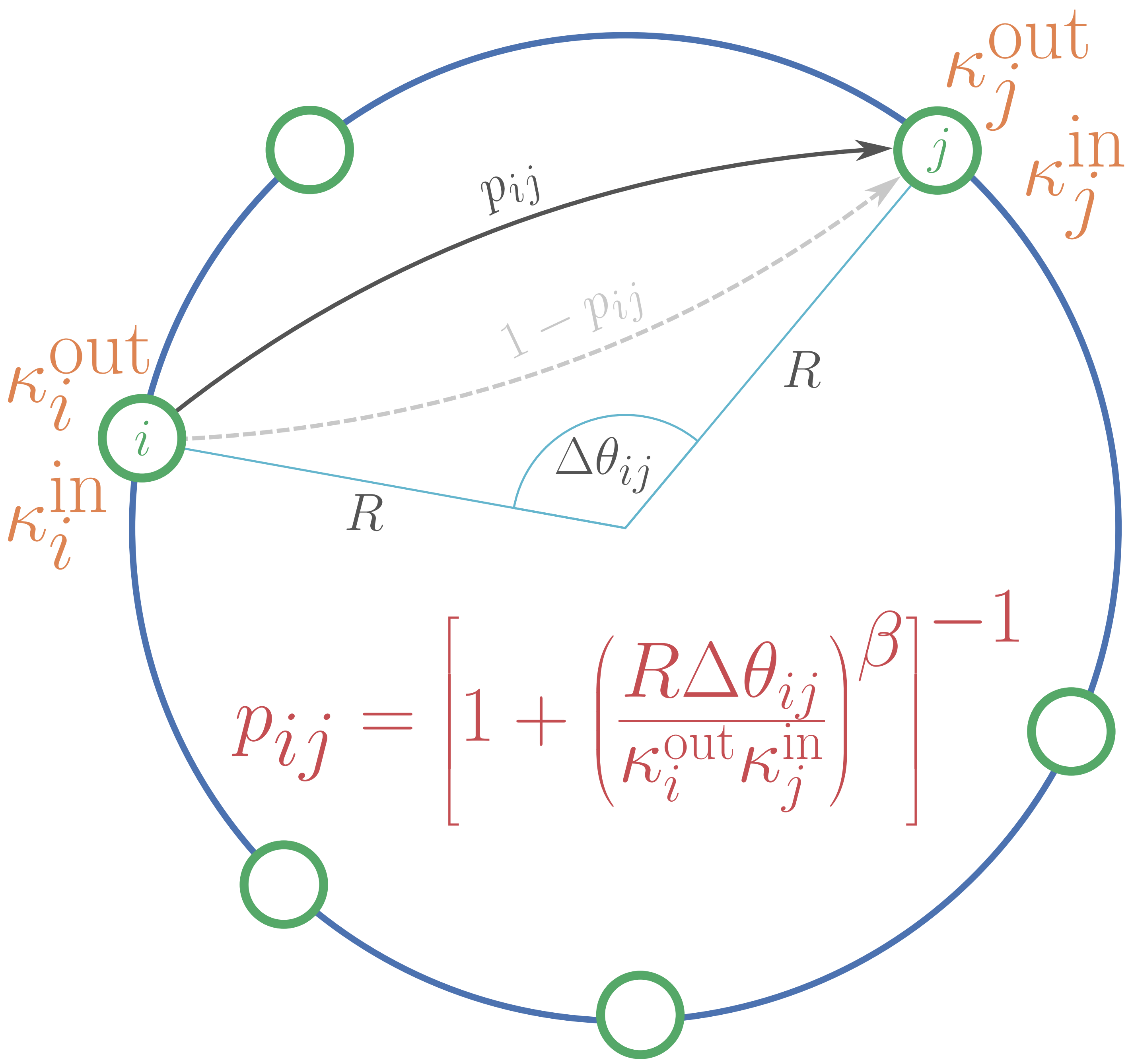
★ fixes the expected in-degree and out-degree of nodes ($\kappa^{in}, \kappa^{out}$) ~> soft directed CM

★ triangle inequality of the underlying metric space -> triangles from pairwise interactions

★ density of triangles tuned with parameter β



The directed $\mathbb{S}^1$ model



$$\mathbb{E}\left[k^{in}|\kappa^{in}\right] \simeq \kappa^{in}$$

$$\mathbb{E}\left[k^{out}|\kappa^{out}\right] \simeq \kappa^{out}$$

$$P(k^{in}, k^{out}) \simeq \iint \frac{[\kappa^{in}]^{k^{in}} e^{-\kappa^{in}}}{k^{in}!} \frac{[\kappa^{out}]^{k^{out}} e^{-\kappa^{out}}}{k^{out}!} \times \rho(\kappa^{in}, \kappa^{out}) d\kappa^{in} d\kappa^{out}$$

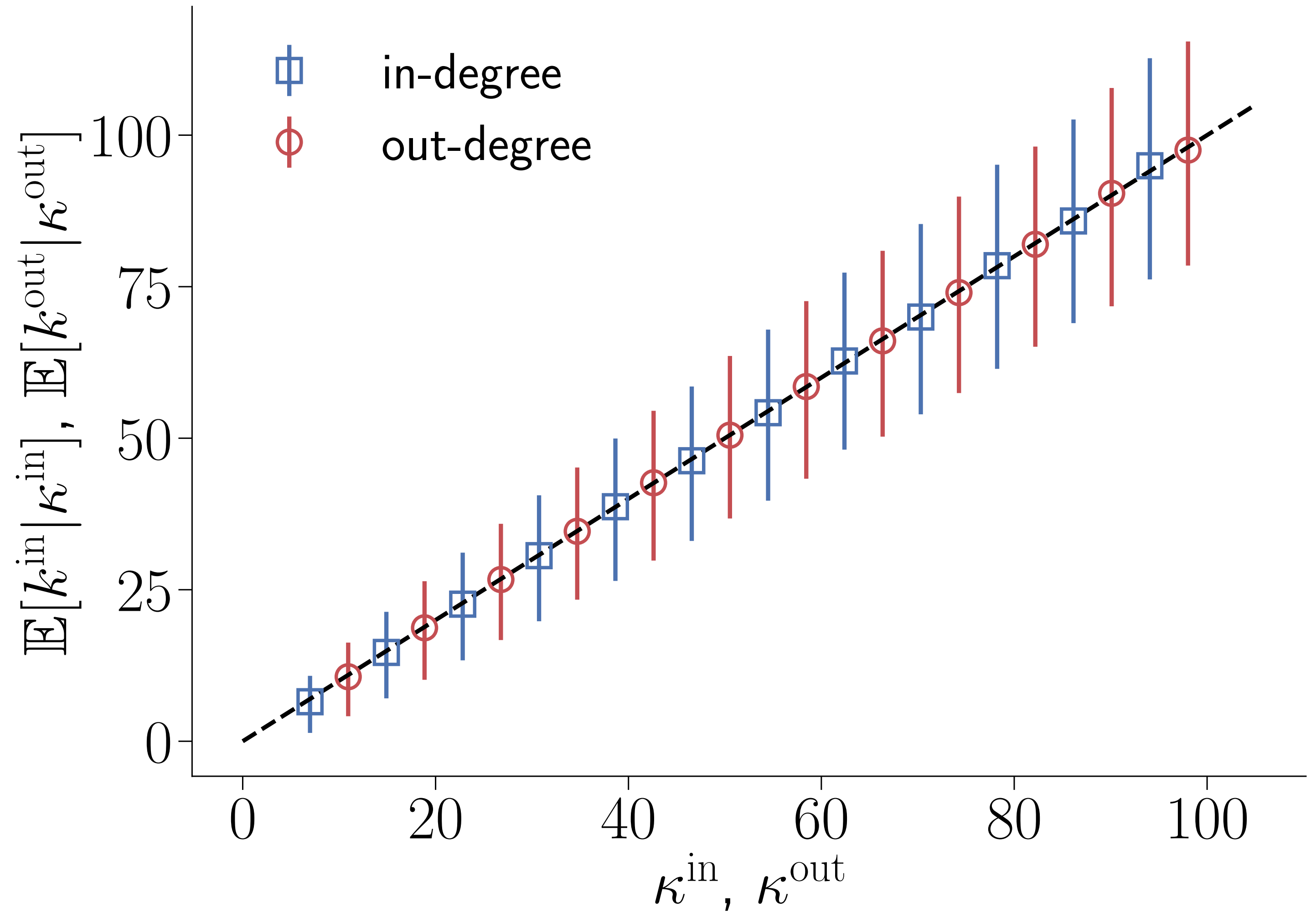
The directed $\mathbb{S}^1$ model

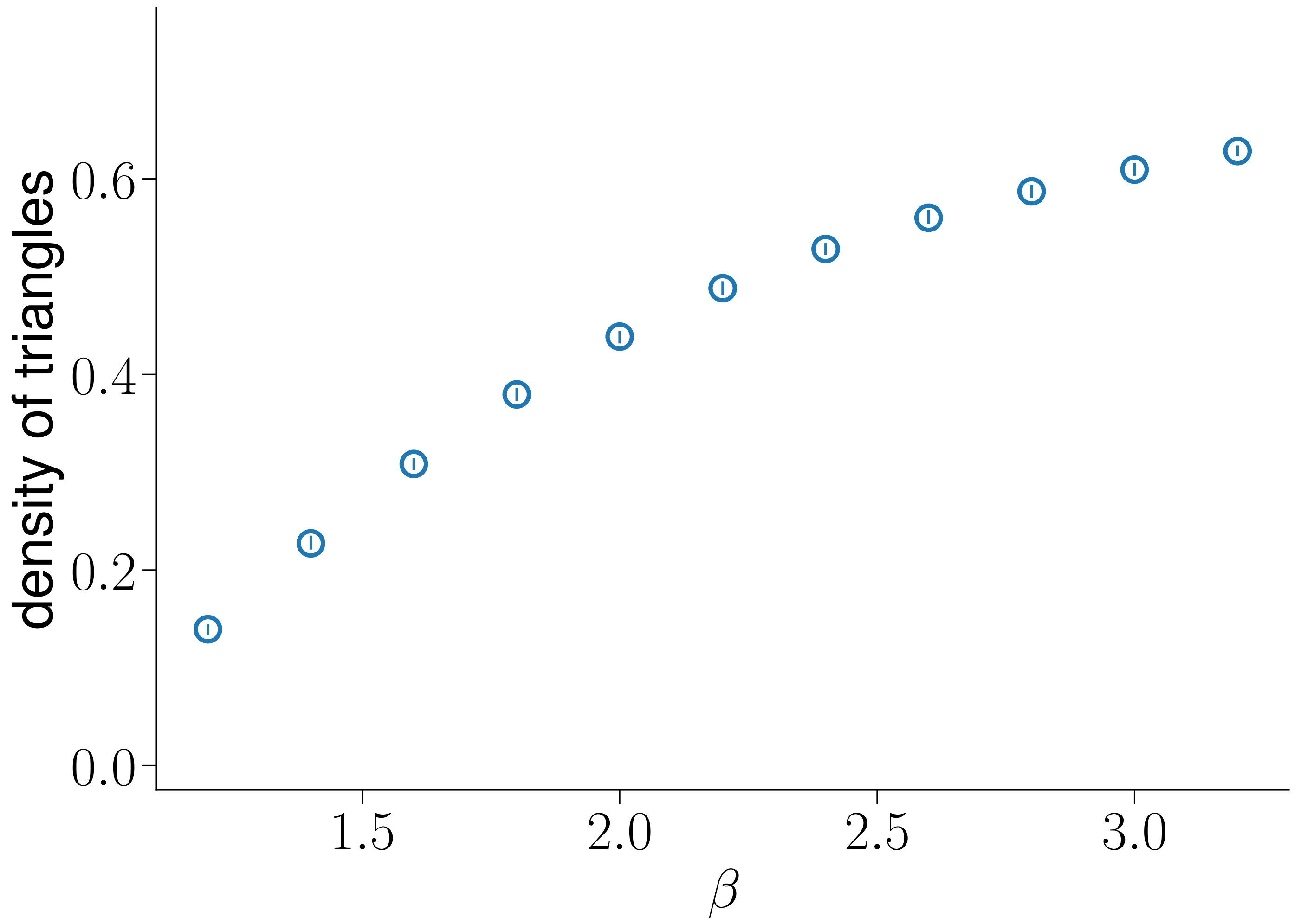
1. Sprinkle N nodes uniformly on a circle of radius R.
2. Assign expected in-degree κ^{in} and out-degree κ^{out} to each node according to some joint pdf $\rho(\kappa^{in}, \kappa^{out})$.
3. Draw a link from node i to node j with probability p_{ij}.

★ fixes the expected in-degree and out-degree of nodes $(\kappa^{in}, \kappa^{out})$ ~> soft directed CM

★ triangle inequality of the underlying metric space -> triangles from pairwise interactions

★ density of triangles tuned with parameter β

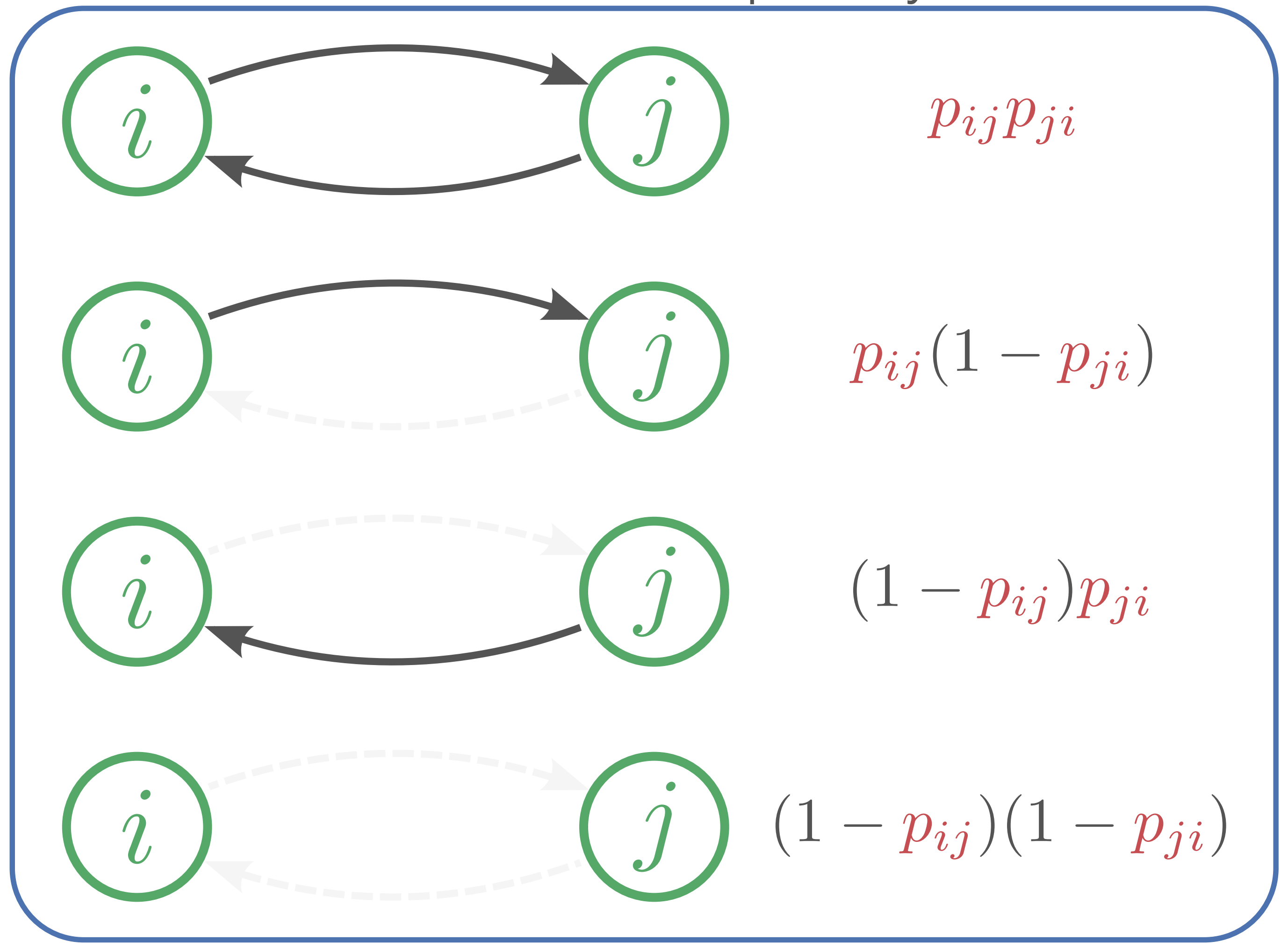




Reciprocity in the directed $\mathbb{S}^1$ model

A reciprocal connection between node i to node j occurs with probability $p_{ij}p_{ji}$.

Fortuitous reciprocity



$$\mathbb{E}\left[r\right] = \mathbb{E}\left[\frac{L^{\leftrightarrow}}{L}\right] = \mathbb{E}\left[\frac{k^{\leftrightarrow}}{k^{\text{out}}}\right] \approx \frac{\mathbb{E}\left[k^{\leftrightarrow}\right]}{\mathbb{E}\left[k^{\text{out}}\right]}$$

$$\simeq \iiiint \frac{\kappa_i^{\text{out}}\kappa_j^{\text{in}}}{\langle\kappa^{\text{in}}\rangle\langle\kappa^{\text{out}}\rangle} \frac{1-\left(\frac{\kappa_i^{\text{out}}}{\kappa_i^{\text{in}}}\frac{\kappa_j^{\text{in}}}{\kappa_j^{\text{out}}}\right)^{\beta-1}}{1-\left(\frac{\kappa_i^{\text{out}}}{\kappa_i^{\text{in}}}\frac{\kappa_j^{\text{in}}}{\kappa_j^{\text{out}}}\right)^{\beta}}$$

$$\times\, \rho(\kappa_i^{\text{in}},\kappa_i^{\text{out}})\rho(\kappa_j^{\text{in}},\kappa_j^{\text{out}})d\kappa_i^{\text{in}}\kappa_i^{\text{out}}d\kappa_j^{\text{in}}\kappa_j^{\text{out}}$$

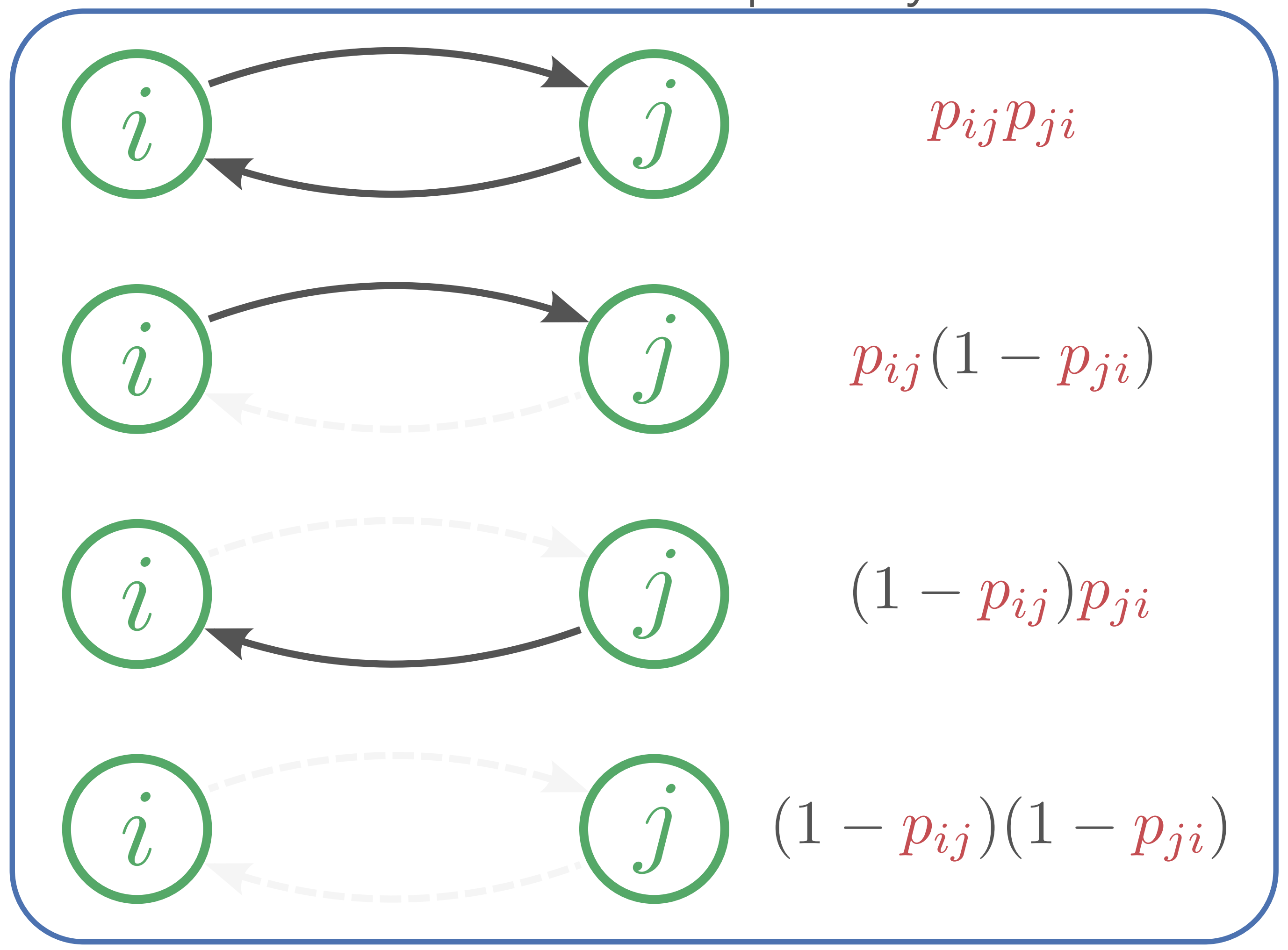
$\kappa^{\text{in}}, \kappa^{\text{out}}$: in-degree and out-degree

β : density of triangles

Reciprocity in the directed $\mathbb{S}^1$ model

A reciprocal connection between node i to node j occurs with probability $p_{ij}p_{ji}$.

Fortuitous reciprocity



$$\mathbb{E}[r] = \mathbb{E}\left[\frac{L^{\leftrightarrow}}{L}\right] = \mathbb{E}\left[\frac{k^{\leftrightarrow}}{k^{\text{out}}}\right] \approx \frac{\mathbb{E}[k^{\leftrightarrow}]}{\mathbb{E}[k^{\text{out}}]}$$

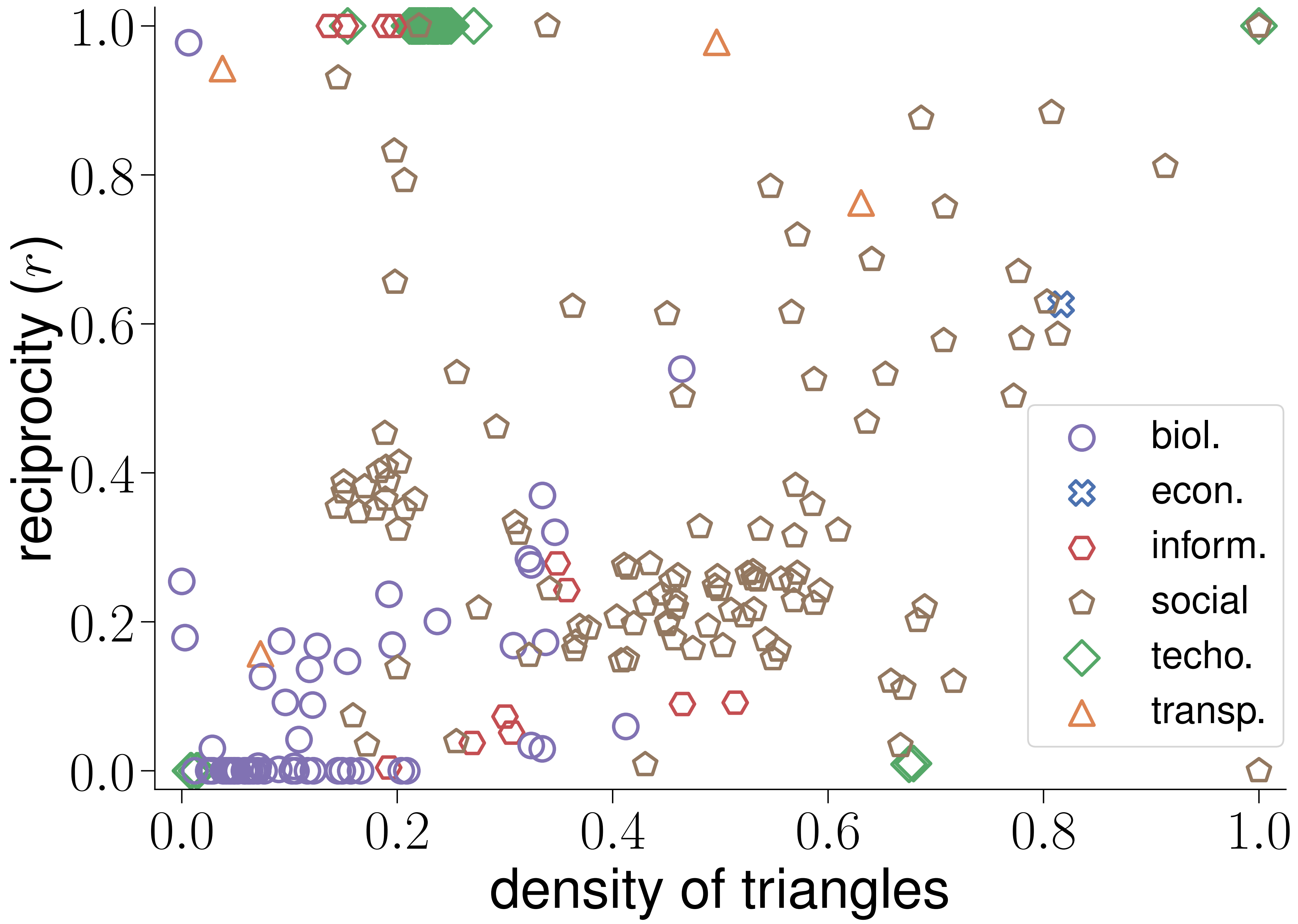
$$\simeq \iiiint \frac{\kappa_i^{\text{out}}\kappa_j^{\text{in}}}{\langle\kappa^{\text{in}}\rangle\langle\kappa^{\text{out}}\rangle} \frac{1-\left(\frac{\kappa_i^{\text{out}}}{\kappa_i^{\text{in}}}\frac{\kappa_j^{\text{in}}}{\kappa_j^{\text{out}}}\right)^{\beta-1}}{1-\left(\frac{\kappa_i^{\text{out}}}{\kappa_i^{\text{in}}}\frac{\kappa_j^{\text{in}}}{\kappa_j^{\text{out}}}\right)^{\beta}}$$

$$\times\, \rho(\kappa_i^{\text{in}},\kappa_i^{\text{out}})\rho(\kappa_j^{\text{in}},\kappa_j^{\text{out}})d\kappa_i^{\text{in}}\kappa_i^{\text{out}}d\kappa_j^{\text{in}}\kappa_j^{\text{out}}$$

$\kappa^{\text{in}}, \kappa^{\text{out}}$: in-degree and out-degree

β : density of triangles

Reciprocity vs. triangles in real directed networks



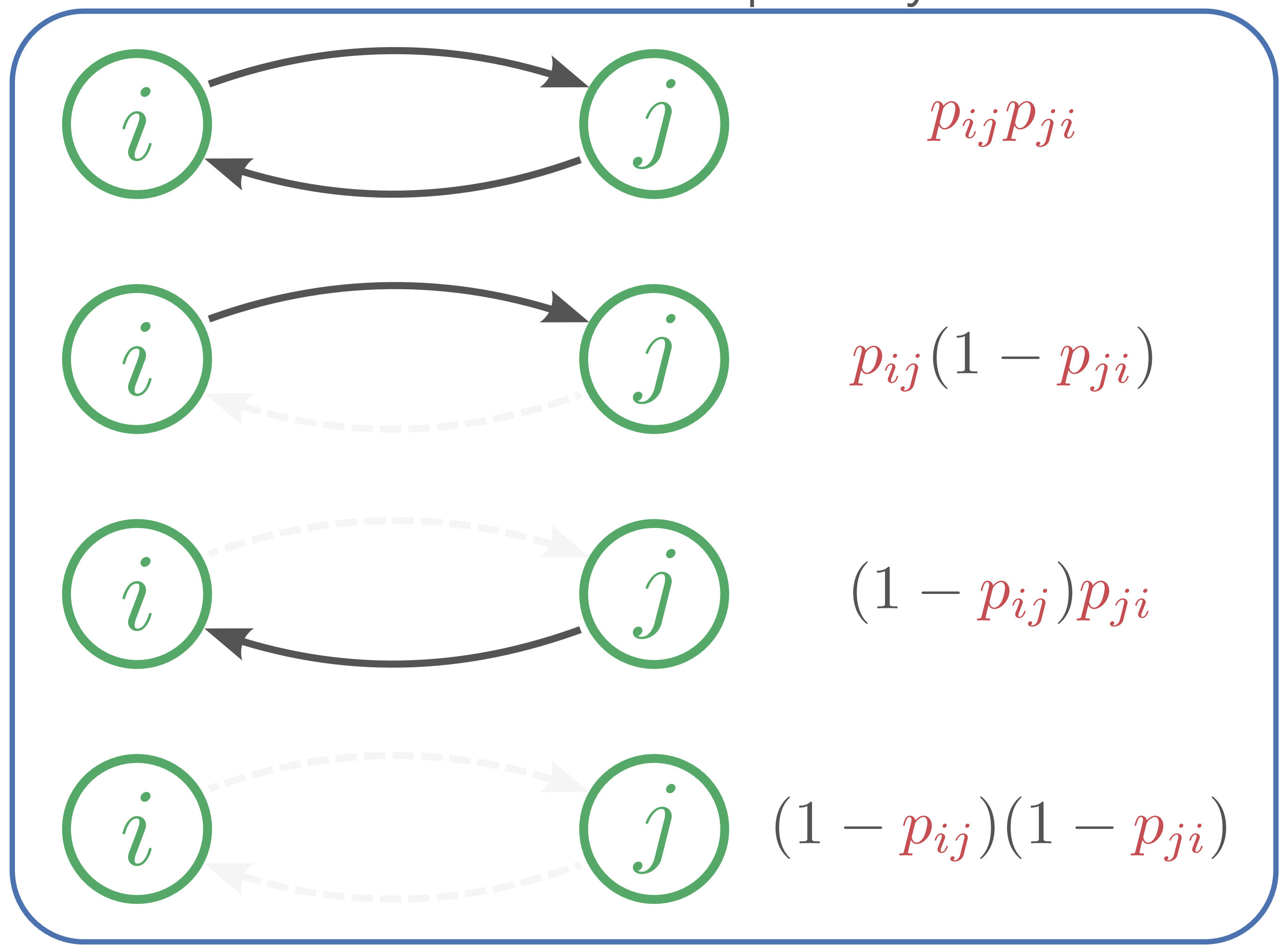
287 network datasets downloaded from Netzschleuder (networks.skewed.de).

A realistic model will need to go beyond fortuitous reciprocity.

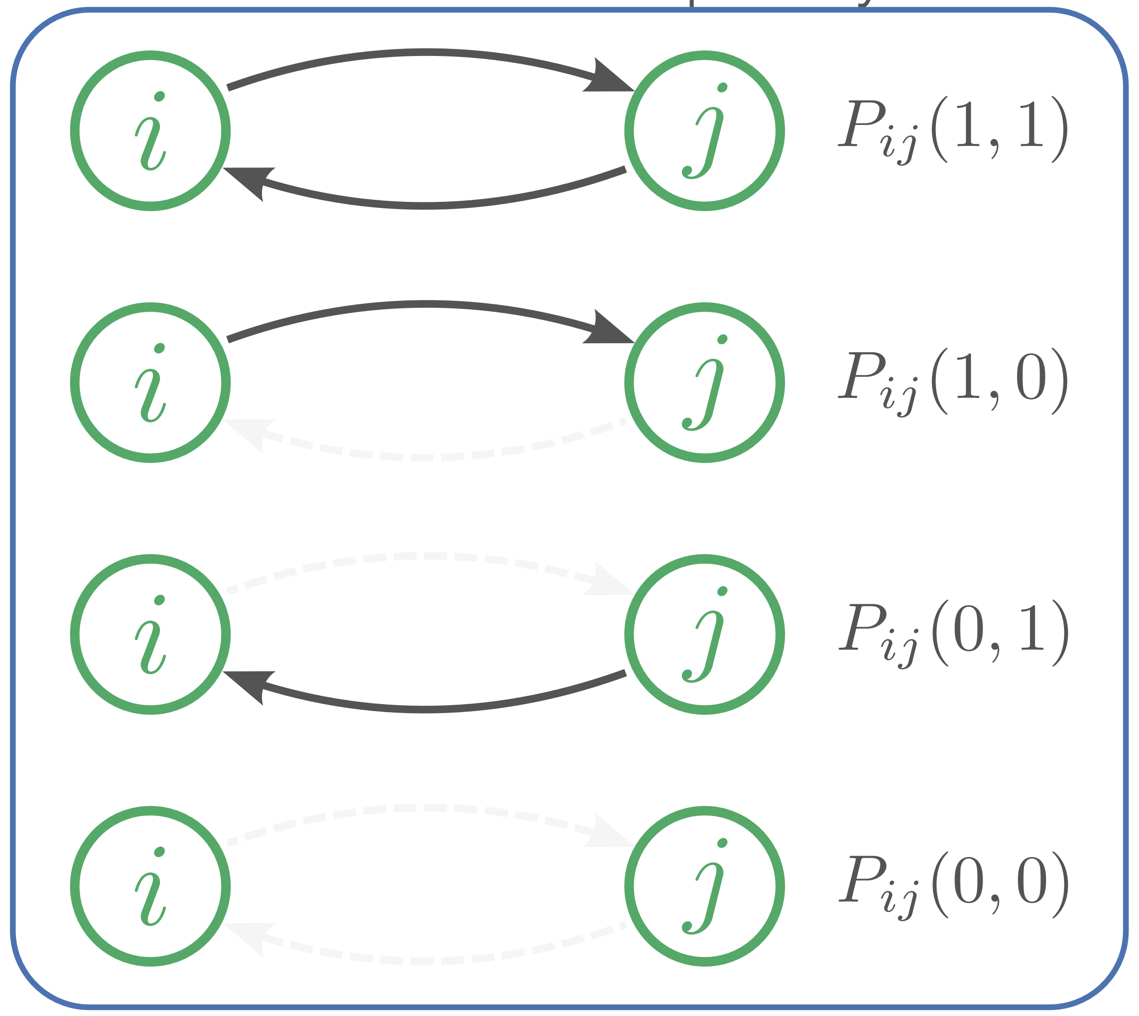
Deliberate reciprocity in random directed networks

A random network model defines the probability p_{ij} for a directed link to exist from node i to node j.

Fortuitous reciprocity

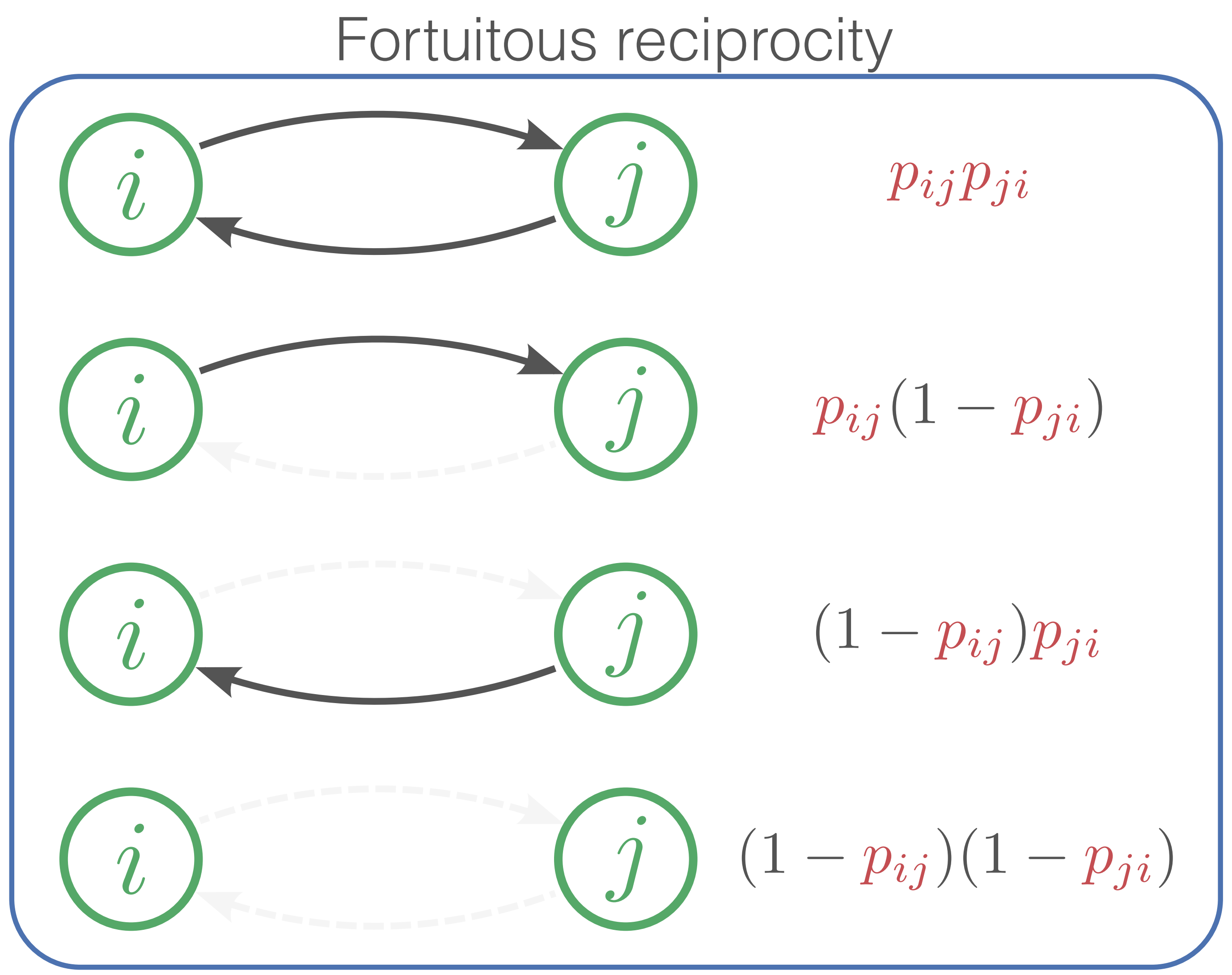


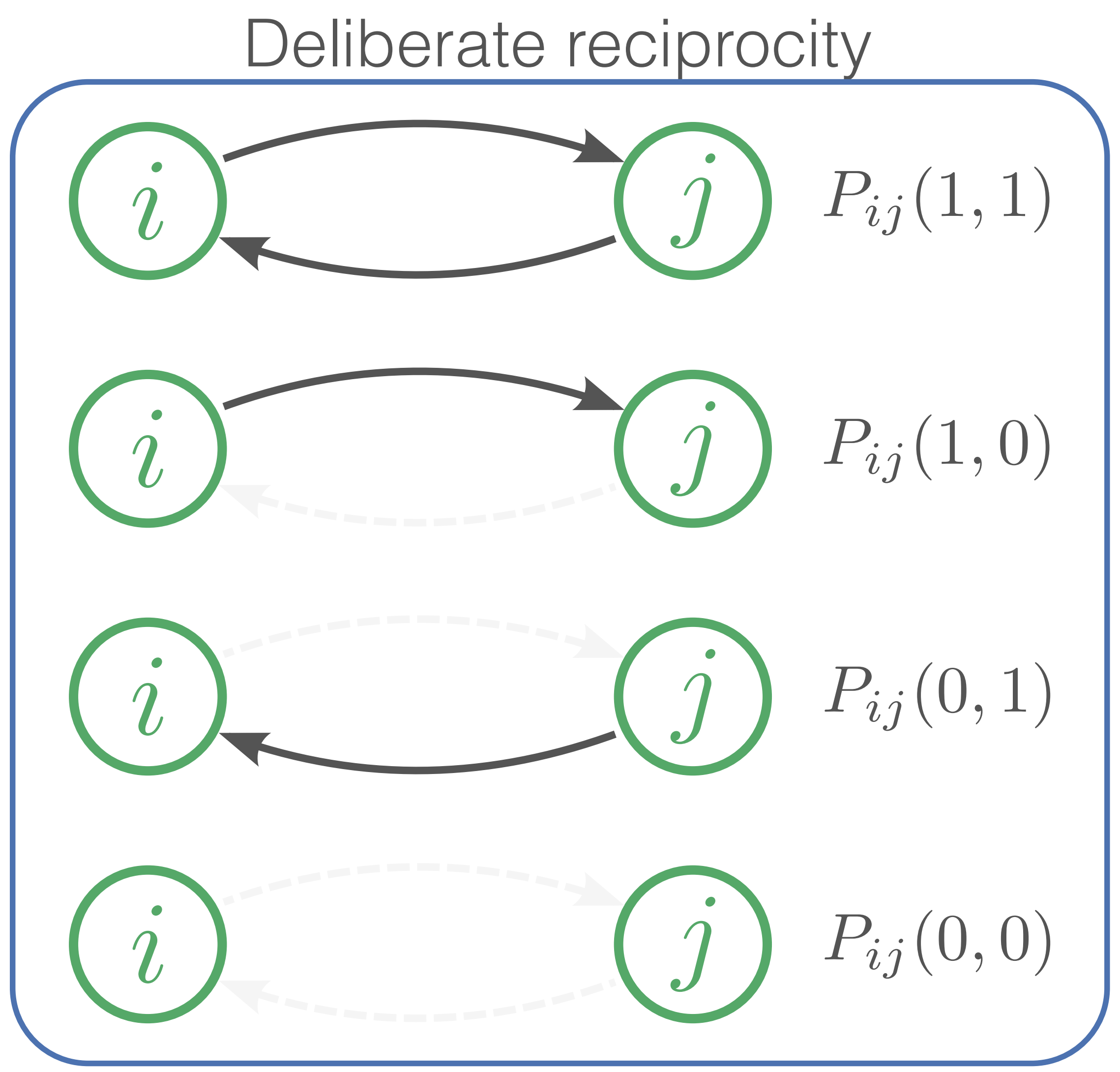
Deliberate reciprocity



Deliberate reciprocity in random directed networks

A random network model defines the probability p_{ij} for a directed link to exist from node i to node j.





Condition 1: Preserves marginal probabilities

$$P_{ij}(1,0) + P_{ij}(1,1) = p_{ij}$$
$$P_{ij}(0,1) + P_{ij}(1,1) = p_{ji}$$

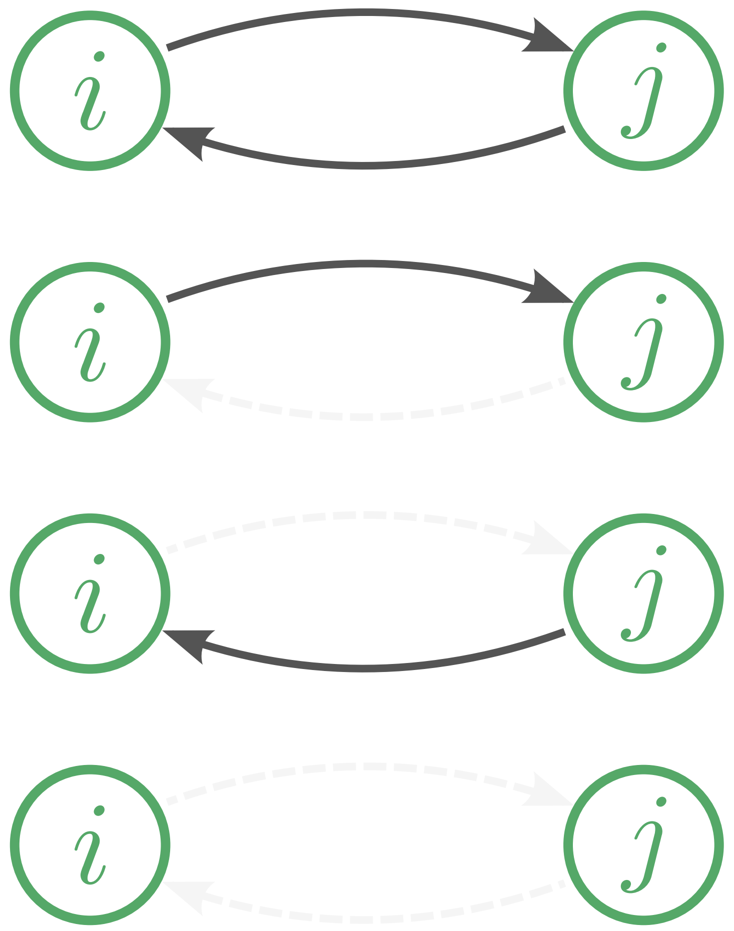
Condition 2: Normalized

$$\sum_{a_{ij}=0}^{1} \sum_{a_{ji}=0}^{1} P_{ij}(a_{ij}, a_{ji}) = 1$$

Deliberate reciprocity in random directed networks

A random network model defines the probability p_{ij} for a directed link to exist from node i to node j.

Fortuitous reciprocity



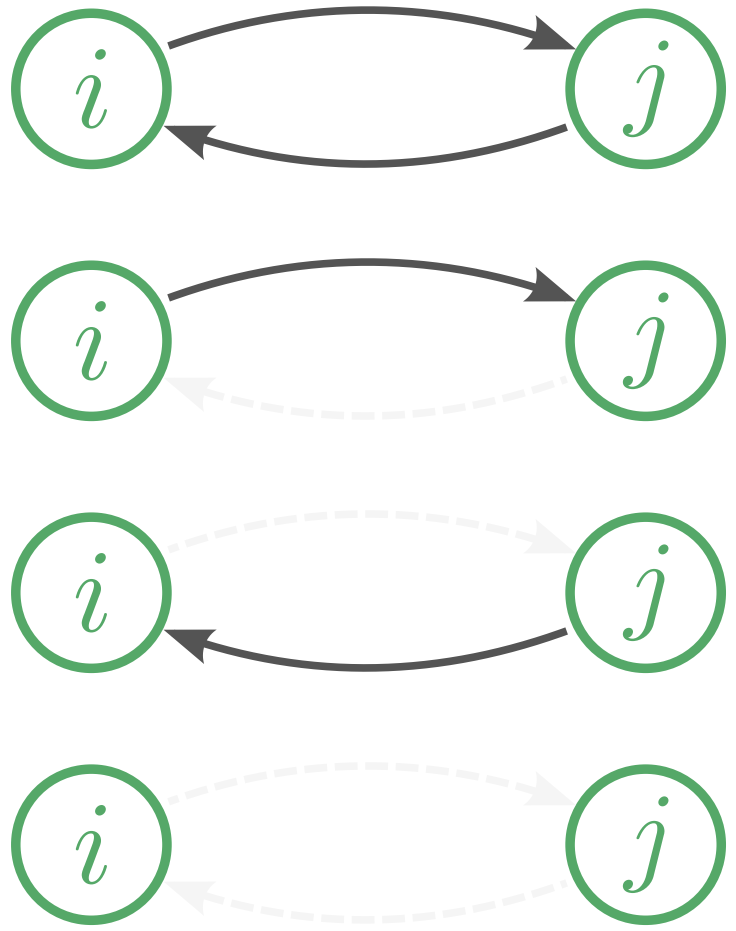
$p_{ij}p_{ji}$

$p_{ij}(1-p_{ji})$

$(1-p_{ij})p_{ji}$

$(1-p_{ij})(1-p_{ji})$

Deliberate reciprocity



$P_{ij}(1,1)$

$P_{ij}(1,0)$

$P_{ij}(0,1)$

$P_{ij}(0,0)$

Condition 1: Preserves marginal probabilities

$$P_{ij}(1,0) + P_{ij}(1,1) = p_{ij}$$
$$P_{ij}(0,1) + P_{ij}(1,1) = p_{ji}$$

Condition 2: Normalized

$$\sum_{a_{ij}=0}^{1} \sum_{a_{ji}=0}^{1} P_{ij}(a_{ij}, a_{ji}) = 1$$

Level of reciprocity controlled with parameter $-1 \leq \nu \leq 1$

$$P_{ij}(1,1) = \begin{cases} (1-\nu)p_{ij}p_{ji} + \nu \min\{p_{ij}, p_{ji}\} & 0 \leq \nu \leq 1 \\ (1+\nu)p_{ij}p_{ji} + \nu(1-p_{ij}-p_{ji})H(p_{ij}+p_{ji}-1) & -1 \leq \nu \leq 0 \end{cases}$$

$\nu = 1$: maximal reciprocity

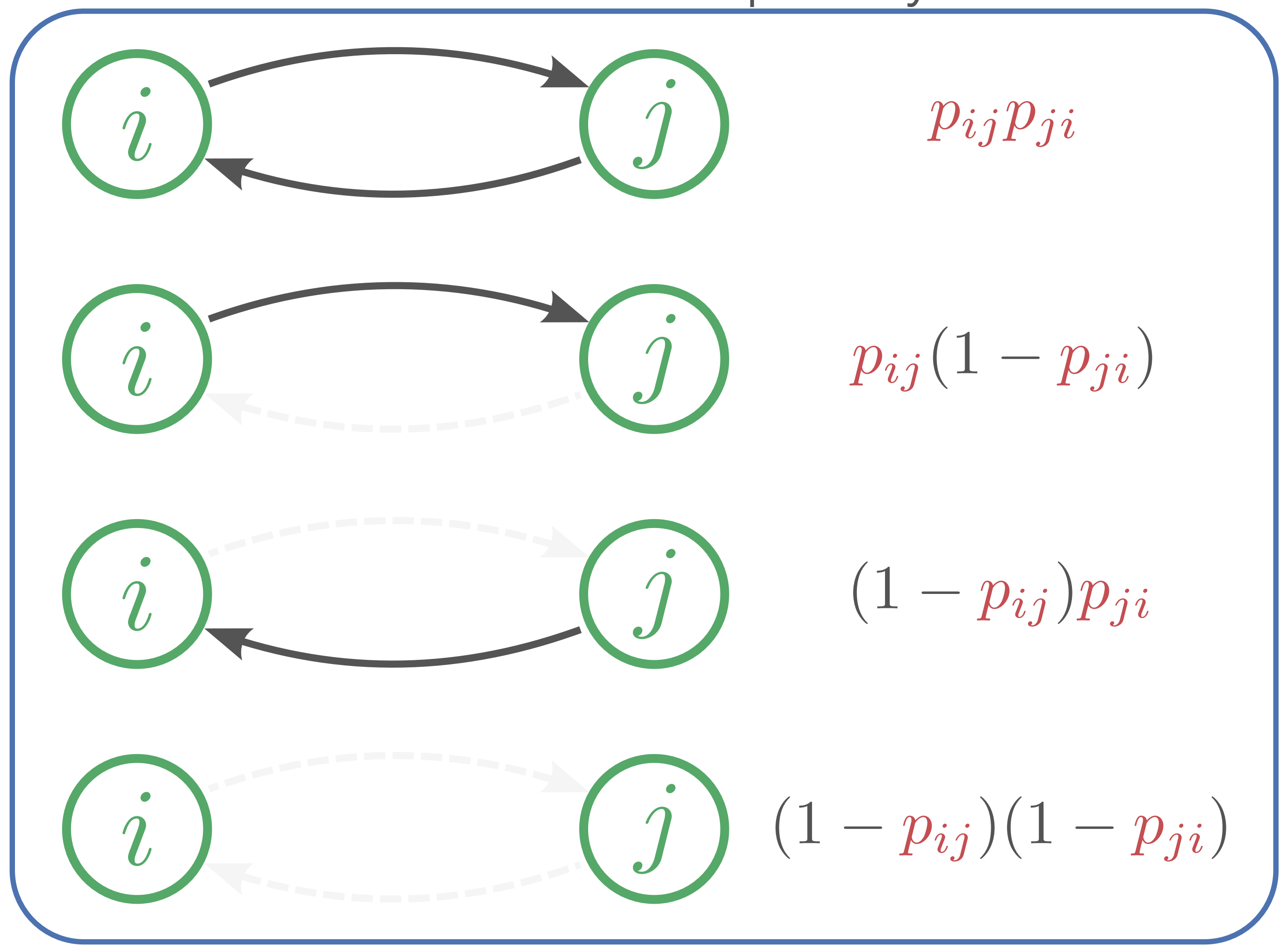
$\nu = 0$: fortuitous reciprocity

$\nu = -1$: minimal reciprocity

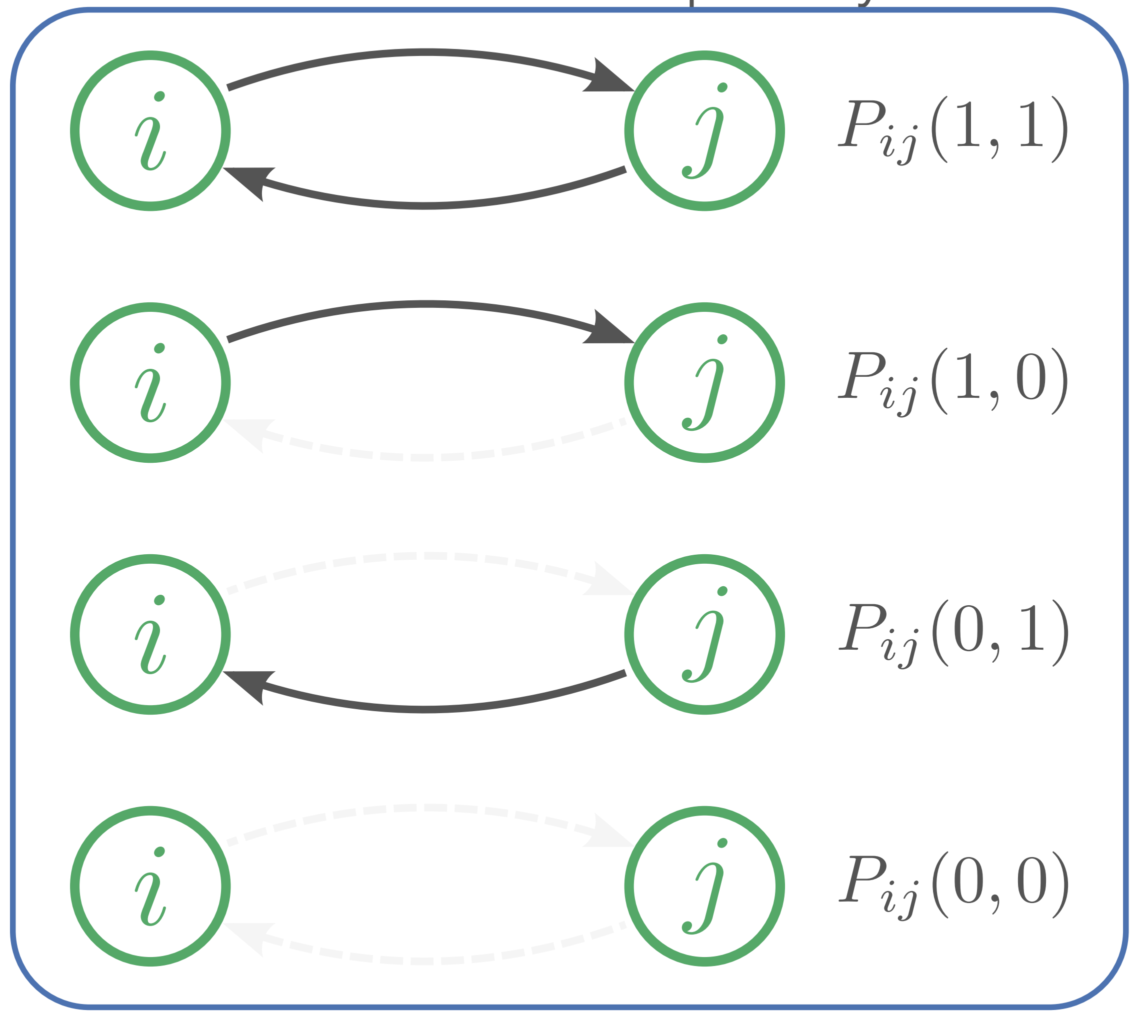
Deliberate reciprocity in random directed networks

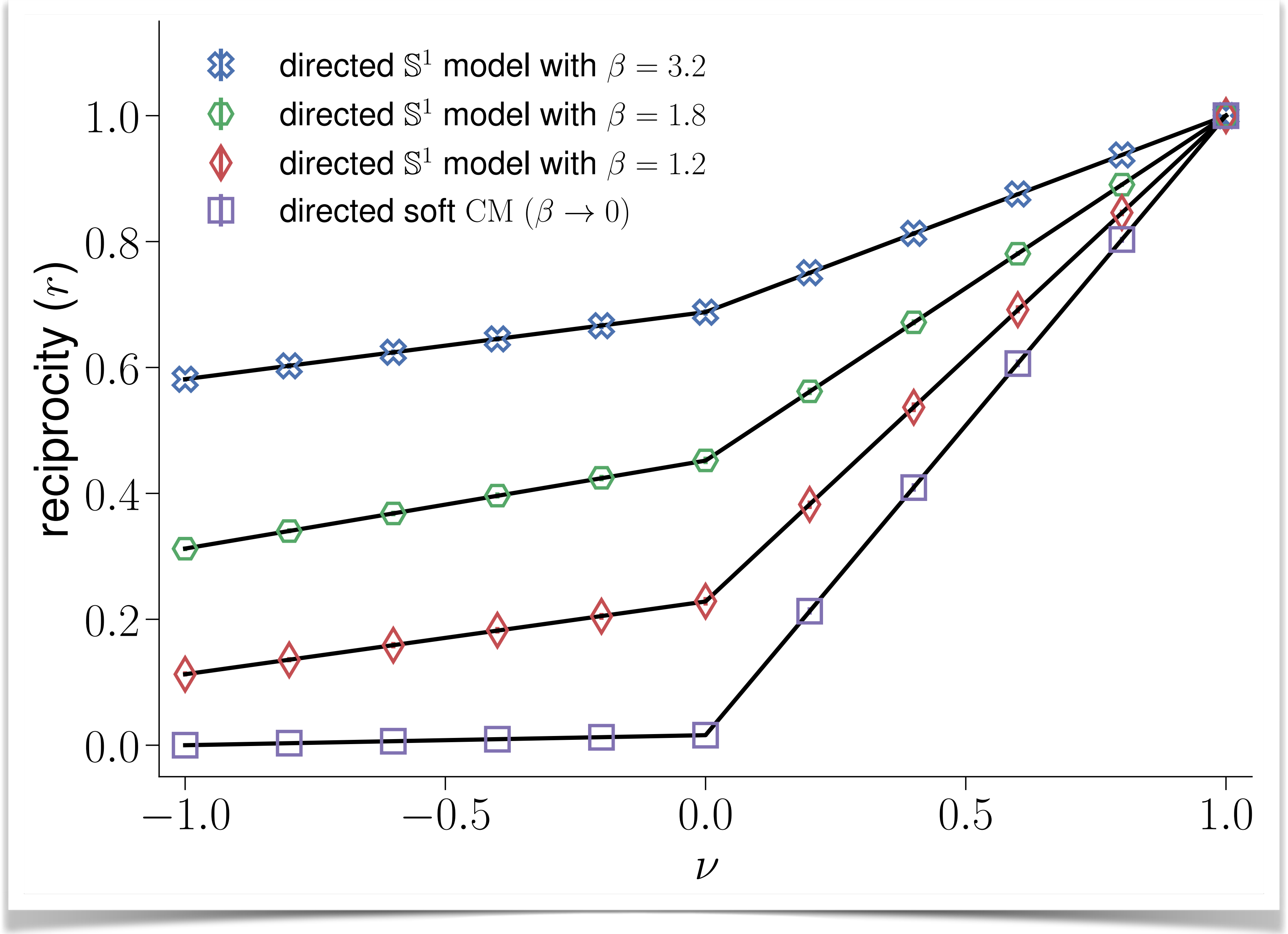
A random network model defines the probability p_{ij} for a directed link to exist from node i to node j.

Fortuitous reciprocity



Deliberate reciprocity





Condition 1: Preserves marginal probabilities

$$P_{ij}(1,0) + P_{ij}(1,1) = p_{ij}$$
$$P_{ij}(0,1) + P_{ij}(1,1) = p_{ji}$$

Condition 2: Normalized

$$\sum_{a_{ij}=0}^{1} \sum_{a_{ji}=0}^{1} P_{ij}(a_{ij}, a_{ji}) = 1$$

Level of reciprocity controlled with parameter $-1 \leq \nu \leq 1$

$$P_{ij}(1,1) = \begin{cases} (1-\nu)p_{ij}p_{ji} + \nu \min\{p_{ij}, p_{ji}\} & 0 \leq \nu \leq 1 \\ (1+\nu)p_{ij}p_{ji} + \nu(1-p_{ij}-p_{ji})H(p_{ij}+p_{ji}-1) & -1 \leq \nu \leq 0 \end{cases}$$

$\nu = 1$: maximal reciprocity
$\nu = 0$: fortuitous reciprocity
$\nu = -1$: minimal reciprocity

Fitting the directed $\mathbb{S}^1$ model to real networks

Inputs from a real network:

1. joint degree distribution $P(k^{\text{in}}, k^{\text{out}})$
2. reciprocity r
3. density of triangles

Assuming uniform angular positions for nodes,

1. infer $\kappa^{\text{in}}, \kappa^{\text{out}}$ to replicate $P(k^{\text{in}}, k^{\text{out}})$ on average (analytical)
2. set ν to reproduce r (analytical)
3. adjust β to recreate the density of triangles (semi-analytical)

Generate a sample of random directed networks:

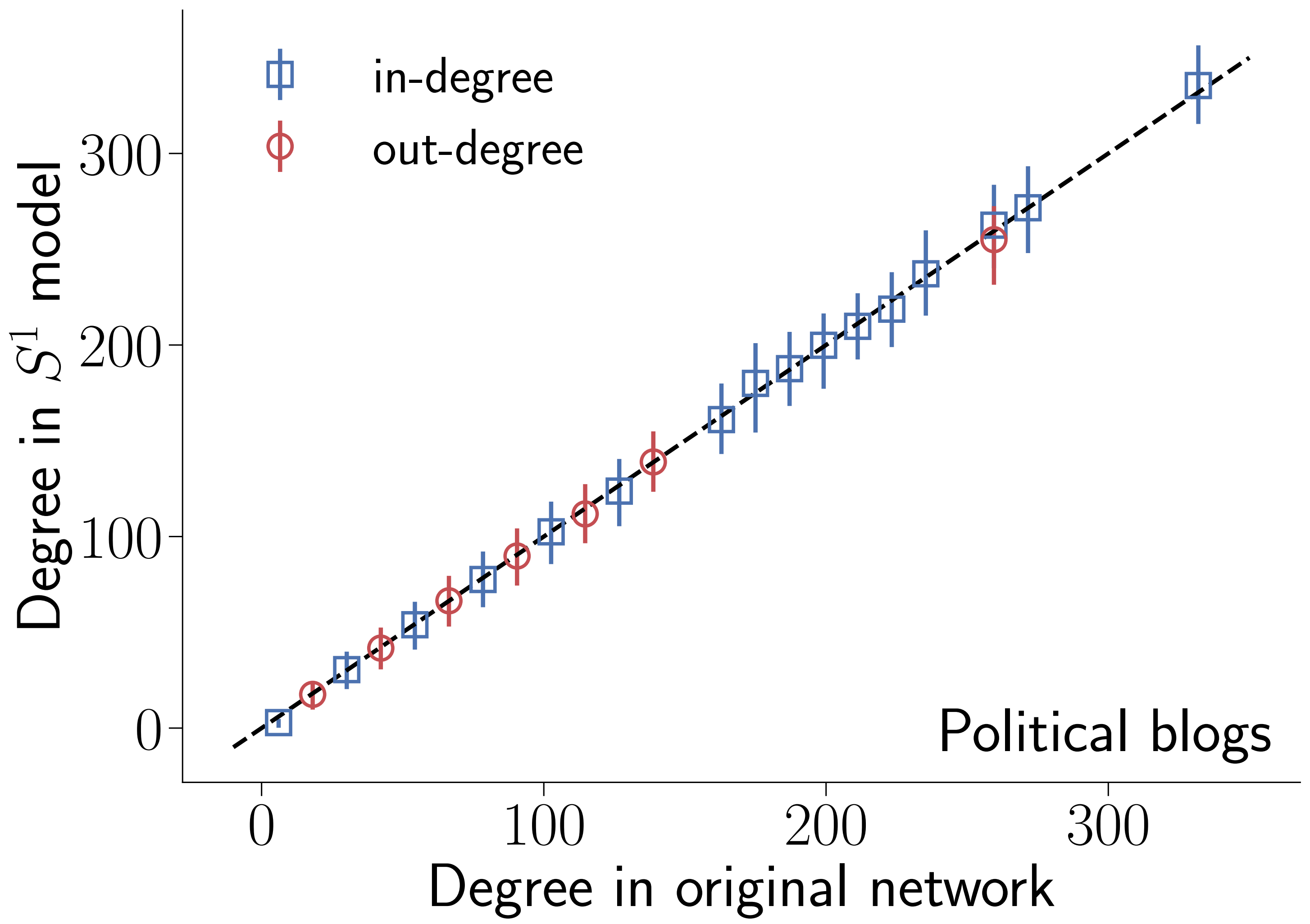
1. assign angular positions randomly
2. draw directed links using the probabilities defined by the framework for deliberate reciprocity

Inference algorithm adapted from New J. Phys. 21, 123033 (2019)

Fitting the directed $\mathbb{S}^1$ model to real networks

Inputs from a real network:

1. joint degree distribution $P(k^{\text{in}}, k^{\text{out}})$
2. reciprocity r
3. density of triangles

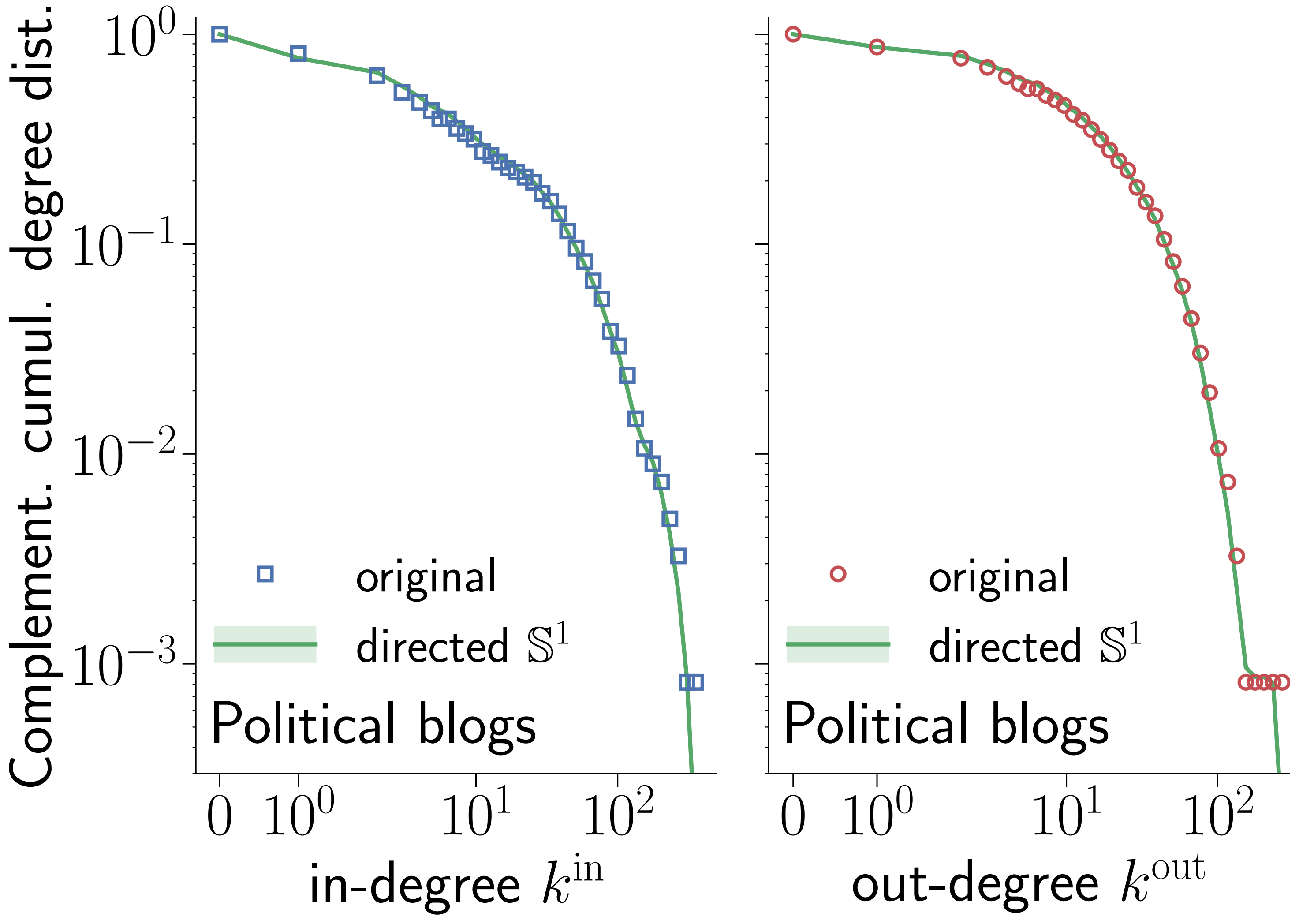


Assuming uniform angular positions for nodes,

1. infer $\kappa^{\text{in}}, \kappa^{\text{out}}$ to replicate $P(k^{\text{in}}, k^{\text{out}})$ on average (analytical)
2. set ν to reproduce r (analytical)
3. adjust β to recreate the density of triangles (semi-analytical)

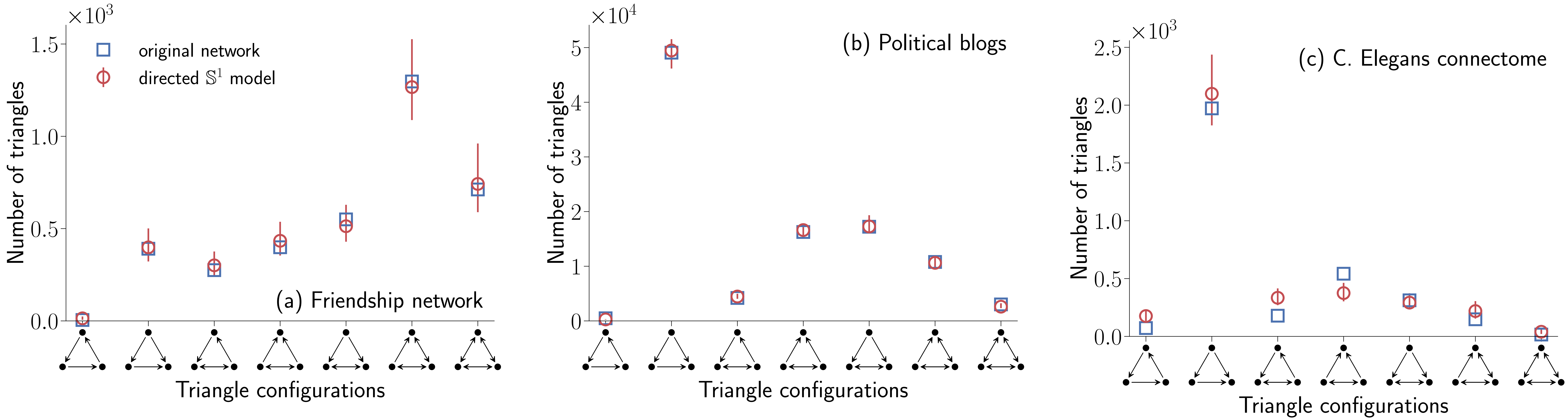
Generate a sample of random directed networks:

1. assign angular positions randomly
2. draw directed links using the probabilities defined by the framework for deliberate reciprocity



Inference algorithm adapted from New J. Phys. 21, 123033 (2019)

Realistic clustering patterns in directed geometric networks



Coupled with an underlying geometry,

1. the joint degree distribution,
2. the reciprocity and
3. the density of triangles

fix the clustering patterns in the network.

Summary

1. Presented a generalization of the $\mathbb{S}^1$ model to directed networks.
2. Proposed a general approach to control reciprocity in any random network model.
3. Showed that the interplay between in/out-degree, reciprocity and clustering in directed networks can be accurately captured by a geometric approach.

Further details

antoineallard.info

@all_are

github.com/networkgeometry/directed-geometric-networks

on arXiv soon

Work done in collaboration with

M. Ángeles Serrano
Universitat de Barcelona
ICREA

Marián Boguñá
Universitat de Barcelona

Sentinelle Nord
UNIVERSITÉ LAVAL
CANADA FIRST RESEARCH EXCELLENCE FUND | APOGÉE CANADA FONDS D'EXCELLENCE EN RECHERCHE
NSERC CRSNG
Fonds de recherche Nature et technologies Québec
compute canada | calcul canada
Calcul Québec