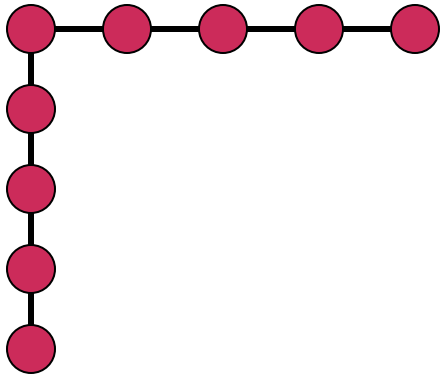
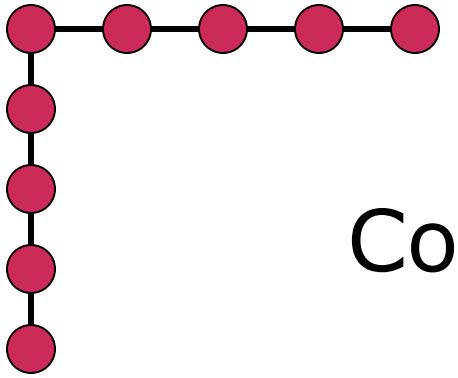


THE UNIVERSITY of WISCONSIN MADISON

Topology Inference from Co-Occurrence Observations

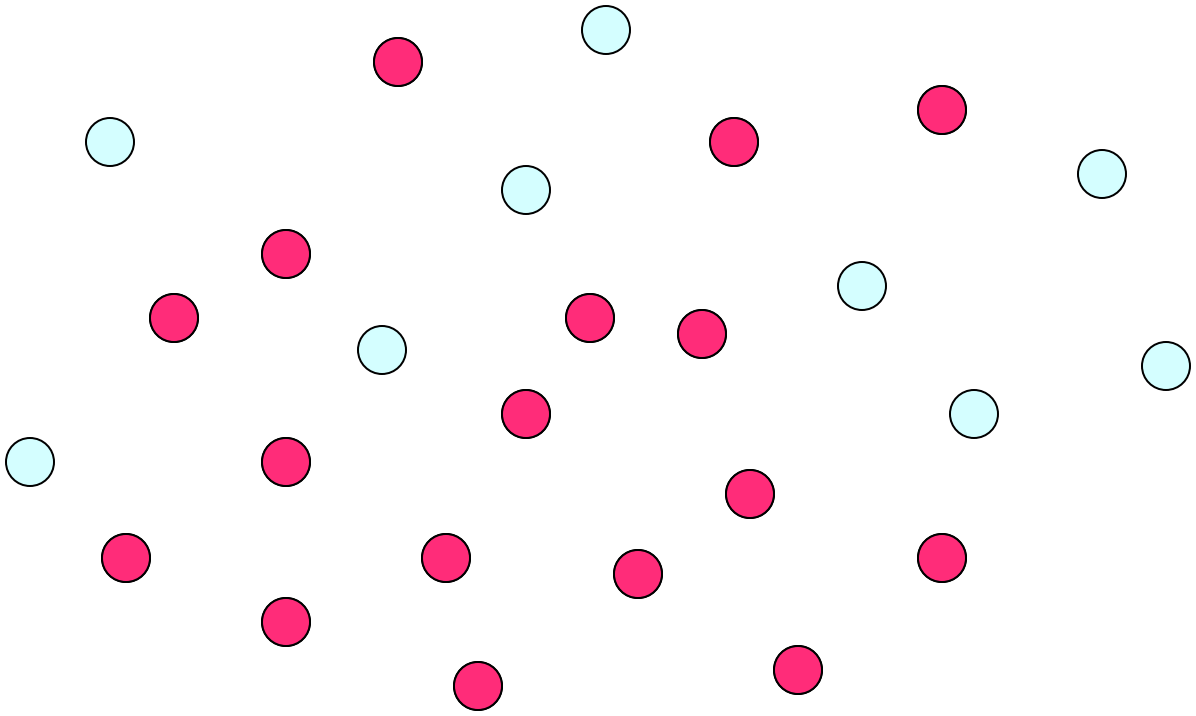
Laura Balzano and Rob Nowak

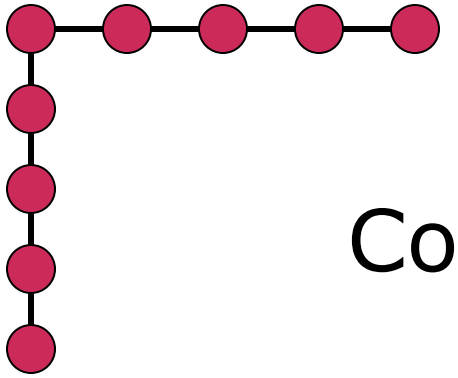
with Michael Rabbat and Matthew Roughan



THE UNIVERSITY of WISCONSIN MADISON

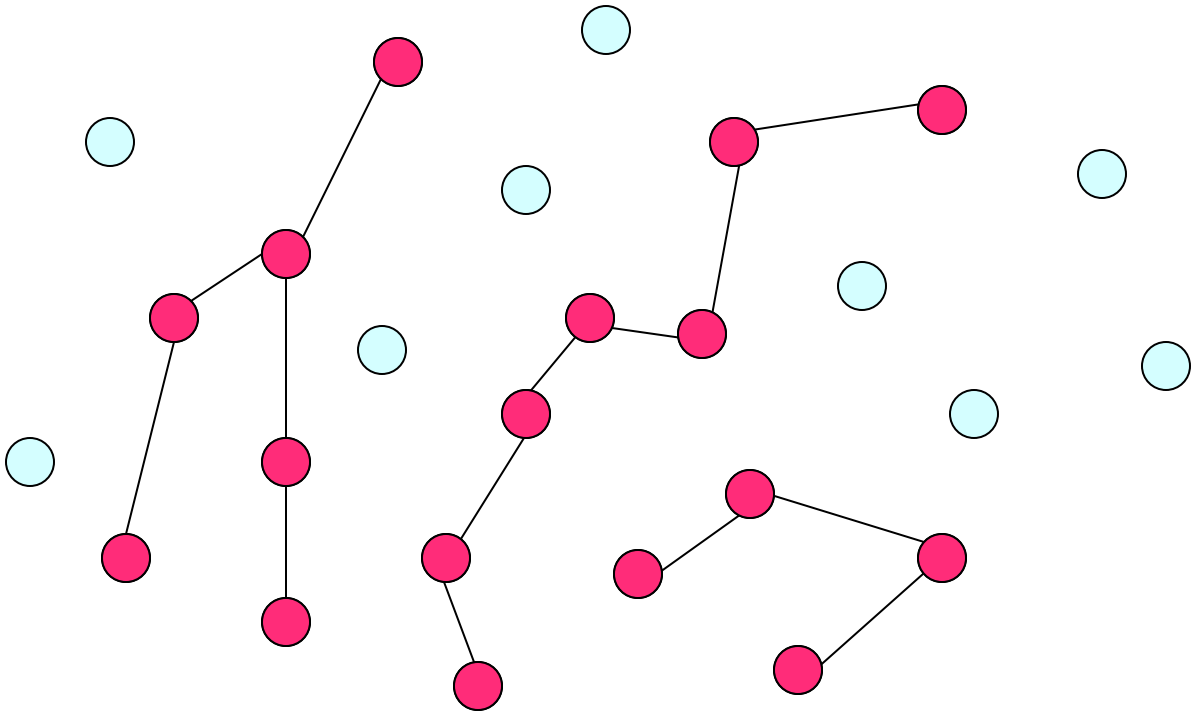
Co-Occurrence observations





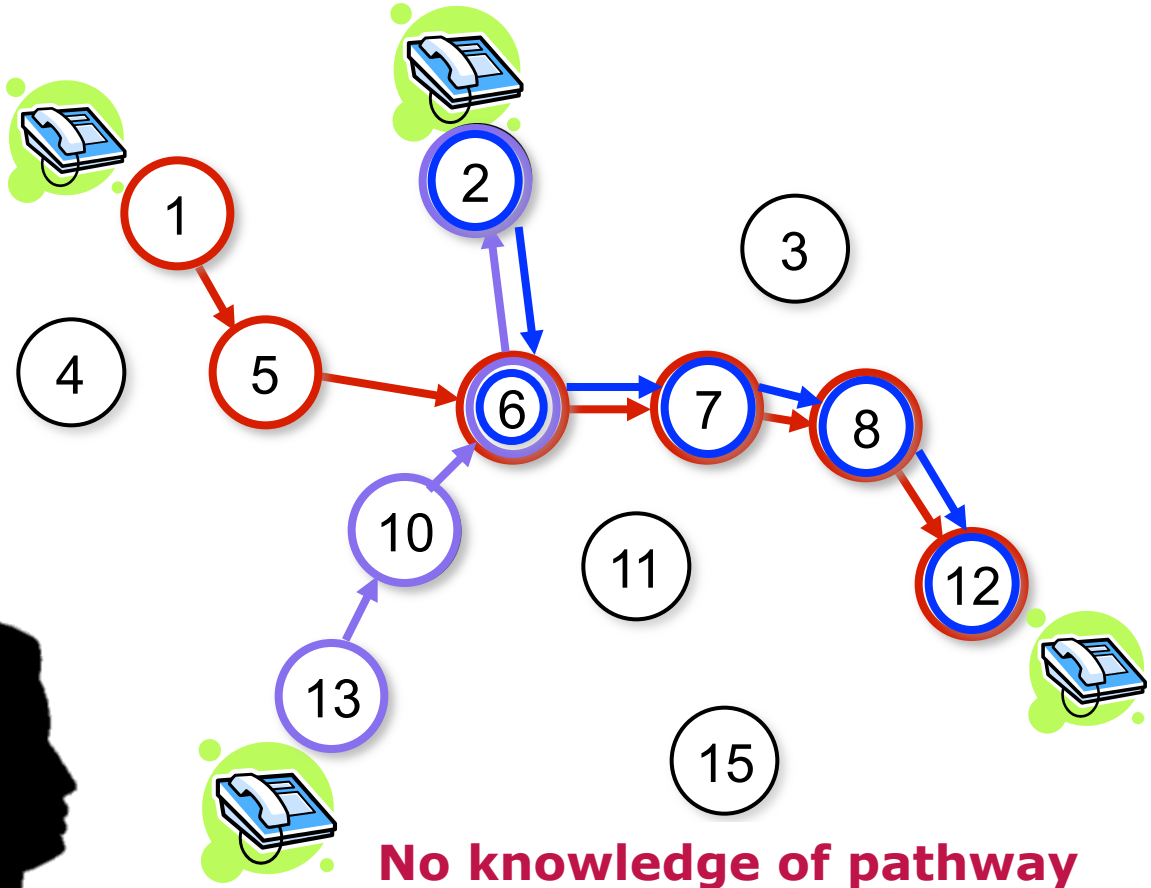
THE UNIVERSITY of WISCONSIN MADISON

Co-Occurrence observations



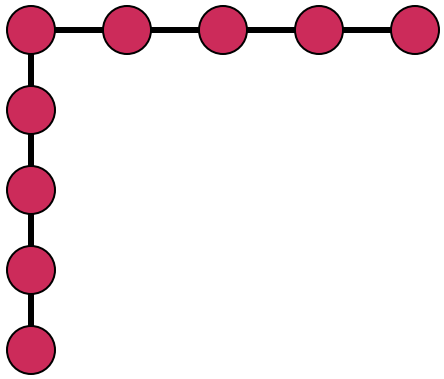
Network Inference for Co-Occurrences (NICO) for telephony

John Treichler



others at 3-letter agency

No knowledge of pathway structures; only know which components are activated in each pathway



THE UNIVERSITY of WISCONSIN MADISON

Network Inference from Linear Pathways

We observe that a set of nodes have been activated.

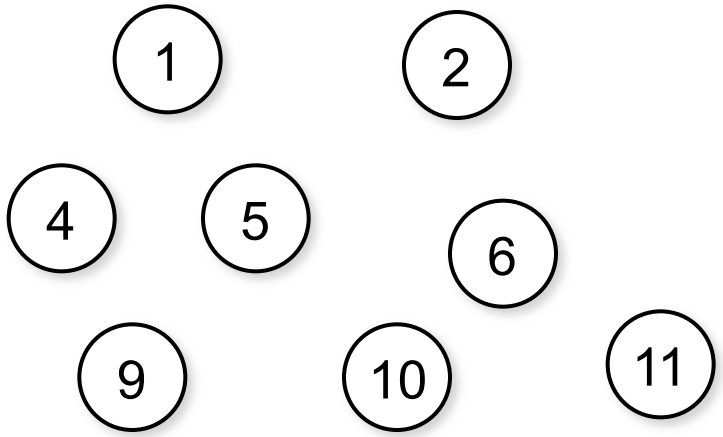
- Assuming the nodes were activated along a linear pathway, we want to reconstruct the pathway through the network.

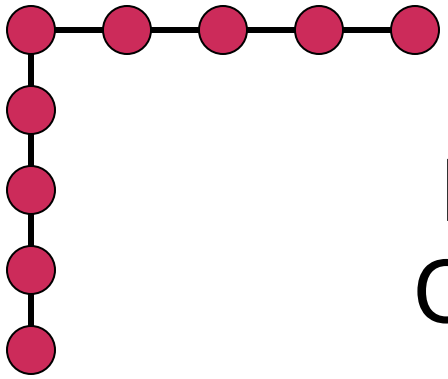
Pathway A: {1, 5, 6, 11}

Pathway B: {4, 5, 6, 10}

Pathway C: {5, 6, 9, 10}

Pathway D: {2, 6, 10}





THE UNIVERSITY of WISCONSIN MADISON

Network Inference from Ordered Linear Pathways

With pathway order information, inference is easy

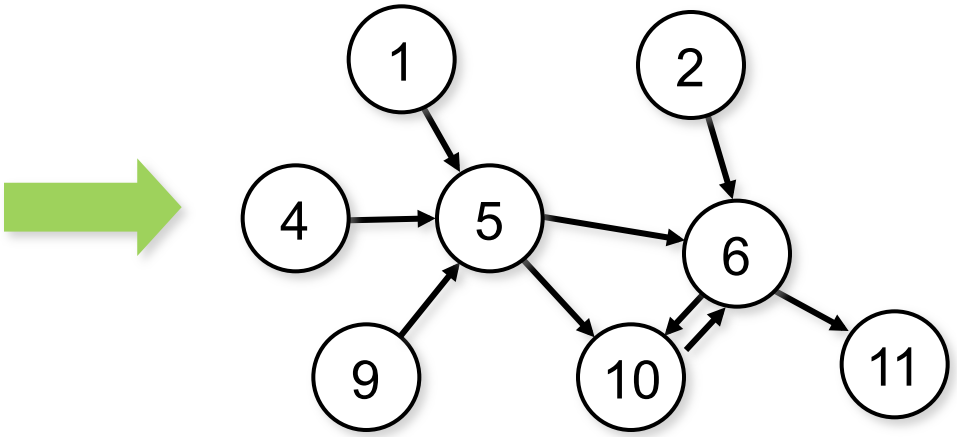
- Insert edges according to paths

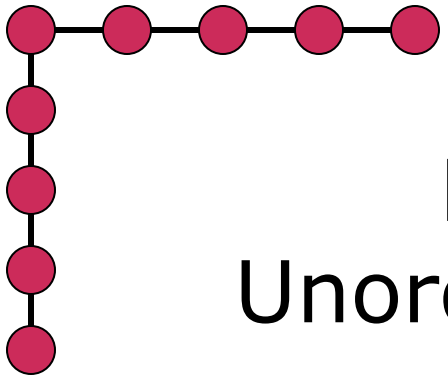
Pathway A: 1 → 5 → 6 → 11

Pathway B: 4 → 5 → 10 → 6

Pathway C: 9 → 5 → 10 → 6

Pathway D: 2 → 6 → 10





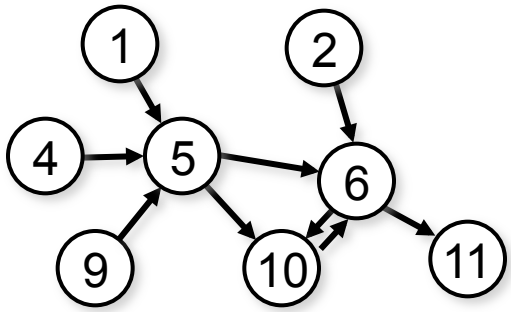
THE UNIVERSITY of WISCONSIN MADISON

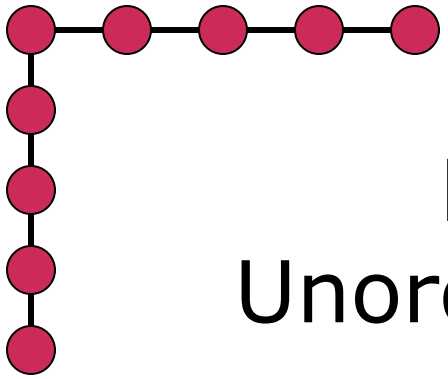
Network Inference from Unordered Linear Co-occurrences

Without pathway order information,

{1,5,6,11}
{4,5,10,6}
{9,5,10,6}
{2,6,10}

1→5→6→11
4→5→10→6
9→5→10→6
2→6→10





THE UNIVERSITY of WISCONSIN MADISON

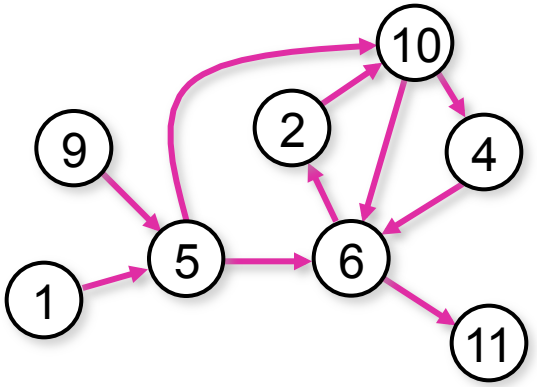
Network Inference from Unordered Linear Co-occurrences

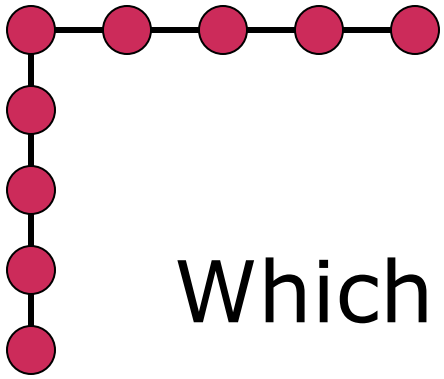
Without pathway order information, every permutation leads to a data-consistent network

- Combinatorial explosion of the solution set

{1,5,6,11}
{4,5,10,6}
{9,5,10,6}
{2,6,10}

1→5→6→11
10→4→6→5
9→5→10→6
6→2→10

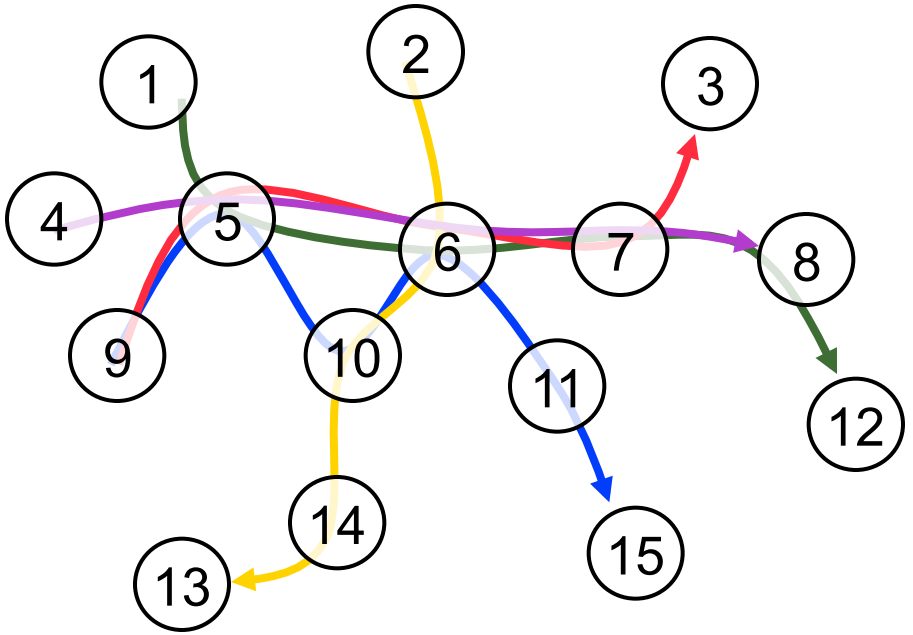




THE UNIVERSITY of WISCONSIN MADISON

Intuition: Which networks are more plausible?

- Vertices that co-occur frequently are probably close together
- Moreover, pathways that are similar (eg **red**, **green**, **purple**) probably have activated vertices arranged in a similar order

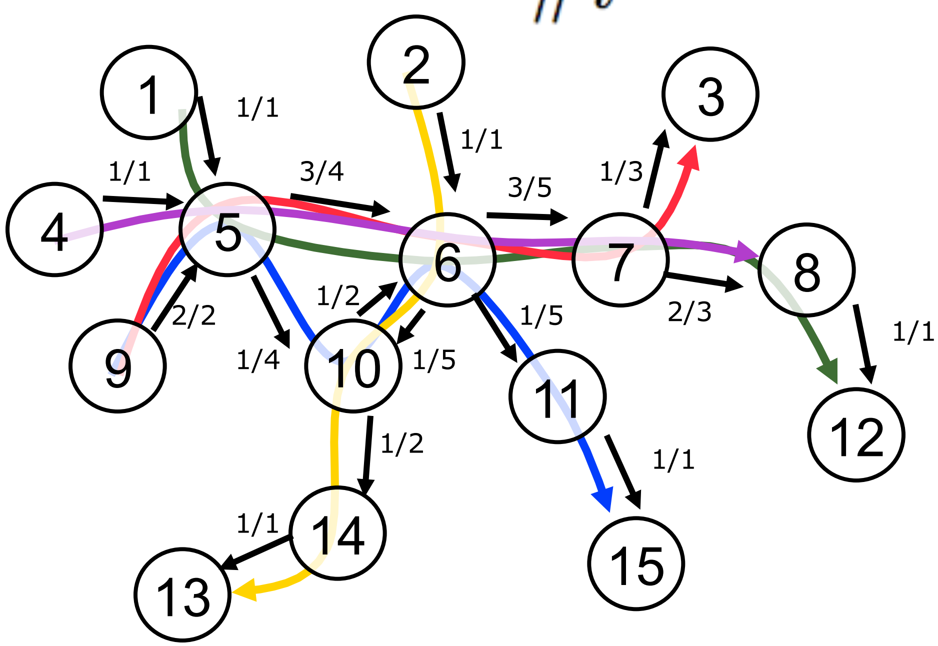


Assumption: The system has been engineered or evolved to re-use existing components (nodes, links) in new pathways.

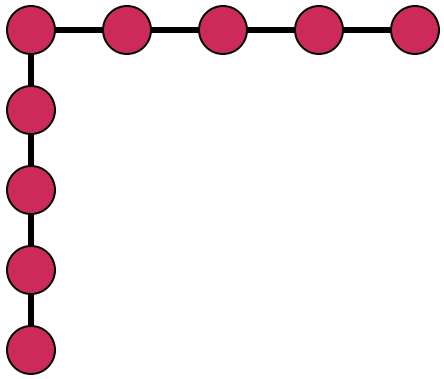
Observation: Signaling probabilities

- Suppose we could directly measure network signaling. Then we could compute empirical next-hop signaling probabilities between nodes:

$$\theta_{i,j} = \frac{\#i \rightarrow j}{\#i}$$



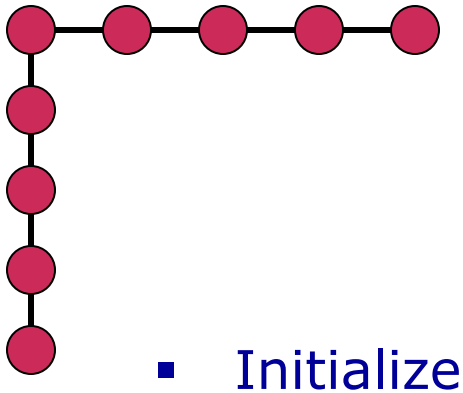
- Then we could calculate a most probable ordering for a particular co-occurrence



THE UNIVERSITY of WISCONSIN MADISON

Estimate them together

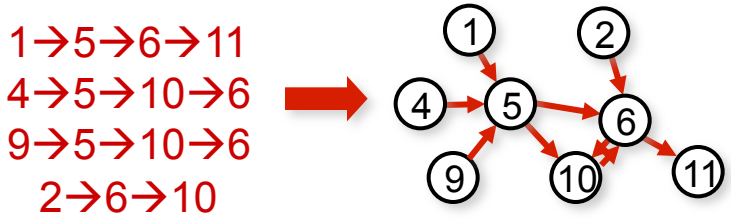
- So if I had the order information, I could estimate the transition probabilities…
- And if I had the transition probabilities, I could estimate the order…

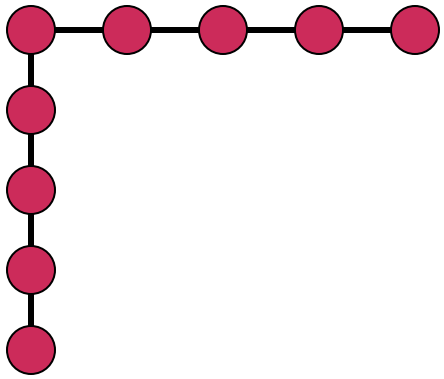


THE UNIVERSITY of WISCONSIN MADISON

EM Algorithm

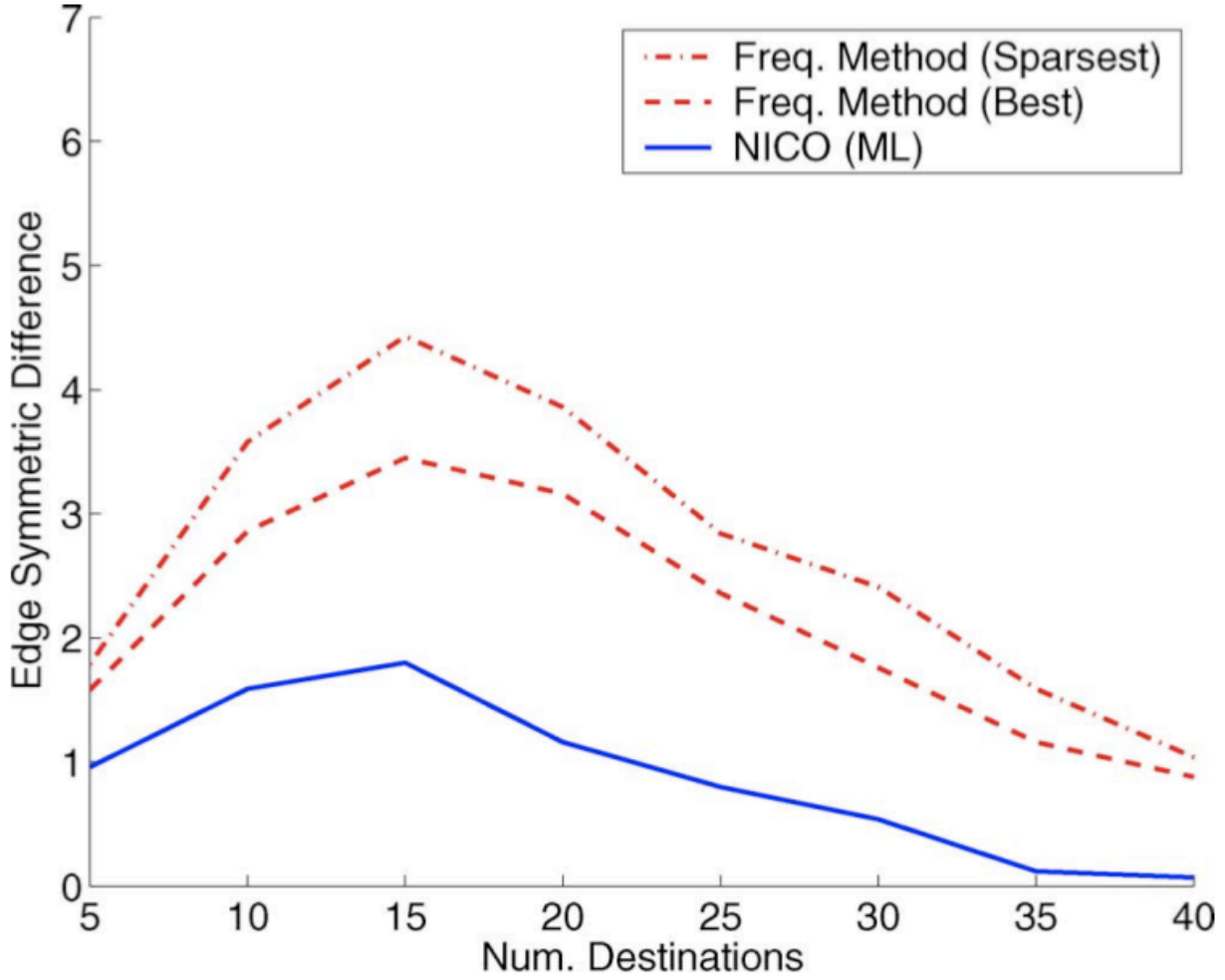
- Initialize Markov chain parameters from unordered co-occurrences
- E-step: Compute likelihood of orders of each pathway using the current estimate of the signaling probabilities
- M-step: Estimate signaling probabilities based on expected permutations.
- When the algorithm converges, use the signaling probabilities and ordered pathways to reconstruct the network

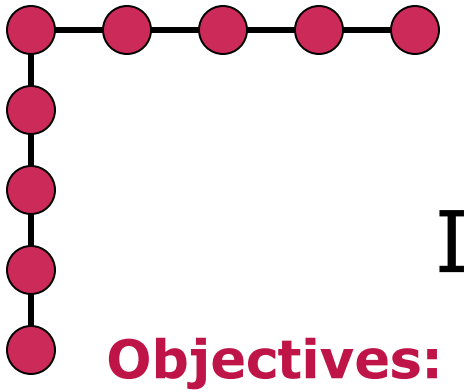




THE UNIVERSITY of WISCONSIN MADISON

Error on Simulated Shortest Path Networks





THE UNIVERSITY of WISCONSIN MADISON

Internet Activation Data

Objectives:

1) `traceroute` measures Internet paths, router activation patterns
2) Real routing, not random walks

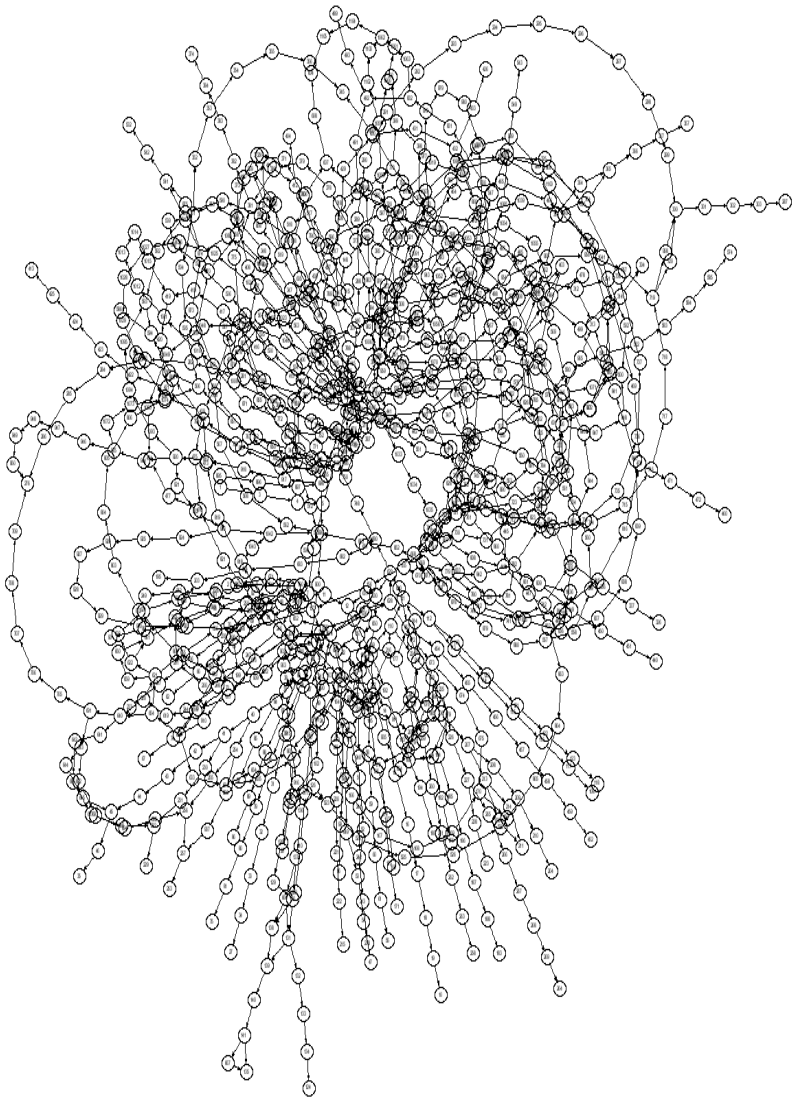
Data Specs:

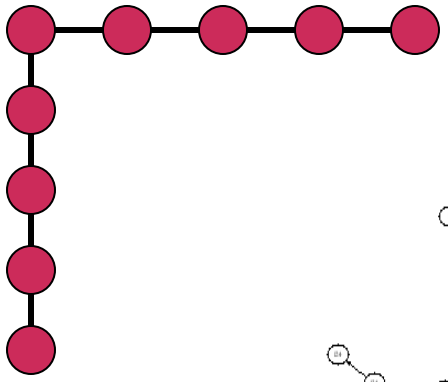
Probe from 3 srcs to 83 dests
1105 node, 1317 edge network
Paths between 8 and 27 hops

Algorithm:

- MCEM when $N_m > 12$
- 50 random initializations

Estimate a 1,105 x 1,105 matrix (with only 1,317 non-zeros) from 3,988 transitions



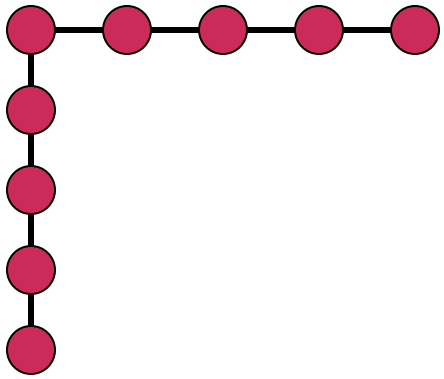


THE UNIVERSITY of WISCONSIN MADISON

Correct (1042)

False positive (273)

False negative (275)



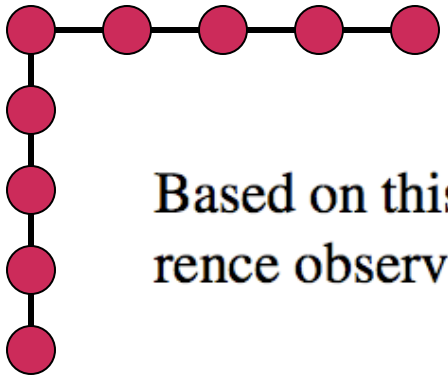
THE UNIVERSITY of WISCONSIN MADISON

Optimization

- Under the Markov chain model, network inference consists in estimating the true transition frequencies θ. Given a prior $P(\theta)$ we can use the maximum a posteriori criterion:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log \mathbb{P}[\mathbf{X}|\boldsymbol{\theta}] + \log \mathbb{P}[\boldsymbol{\theta}]$$

- If $P(\theta)$ is uniform over all the possible θ then this reduces to the maximum likelihood criterion.



Based on this model, we can write the likelihood of a co-occurrence observation $\mathbf{x}$ conditioned on the permutation π as

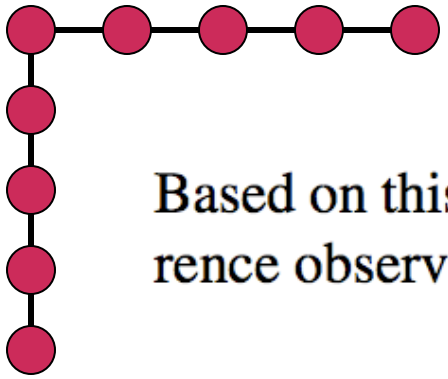
$$\mathbb{P}[\mathbf{x}|\pi,\boldsymbol{\theta}] = \theta_{0,x_{\pi(1)}} \prod_{t=2}^{T} \theta_{x_{\pi(t-1)},x_{\pi(t)}}. \tag{2}$$

Since $\mathbb{P}[\pi] = 1/(T!)$, for all $\pi \in \mathbb{S}_T$, marginalization over all permutations leads to

$$\mathbb{P}[\mathbf{x}|\boldsymbol{\theta}] = \frac{1}{T!} \sum_{\pi \in \mathbb{S}_T} \mathbb{P}[\mathbf{x}|\pi,\boldsymbol{\theta}]. \tag{3}$$

Finally, assuming that co-occurrence observations are independent, and taking the logarithm, gives

$$\log \mathbb{P}[\mathbf{X}|\boldsymbol{\theta}] = \sum_{n=1}^{N} \left[\log \left(\sum_{\pi \in \mathbb{S}_{T_n}} \mathbb{P}[\mathbf{x}^{(n)}|\pi,\boldsymbol{\theta}] \right) - \log(T_n!) \right]. \tag{4}$$



Based on this model, we can write the likelihood of a co-occurrence observation $\mathbf{x}$ conditioned on the permutation π as

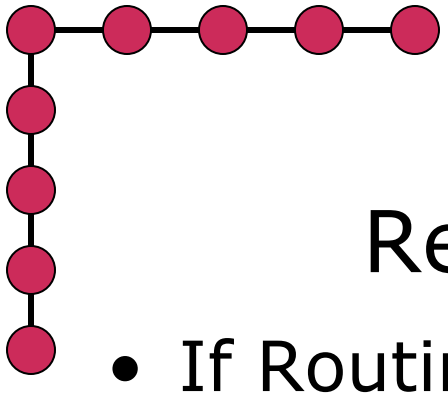
$$\mathbb{P}[\mathbf{x}|\pi,\boldsymbol{\theta}] = \theta_{0,x_{\pi(1)}} \prod_{t=2}^{T} \theta_{x_{\pi(t-1)},x_{\pi(t)}}. \tag{2}$$

Since $\mathbb{P}[\pi] = 1/(T!)$ for all $\pi \in \mathbb{S}_T$, marginalization over all permutations π gives

$$\mathbb{P}[\mathbf{x}|\boldsymbol{\theta}] = \frac{1}{T!} \sum_{\pi \in \mathbb{S}_T} \mathbb{P}[\mathbf{x}|\pi,\boldsymbol{\theta}]. \tag{3}$$

Finally, assuming that co-occurrence observations are independent, and taking the logarithm, gives

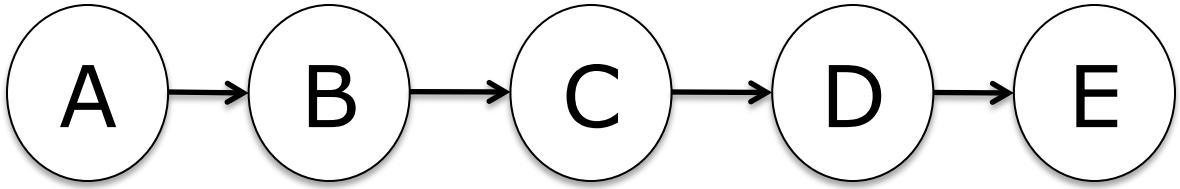
$$\log \mathbb{P}[\mathbf{X}|\boldsymbol{\theta}] = \sum_{n=1}^{N} \left[\log \left(\sum_{\pi \in \mathbb{S}_{T_n}} \mathbb{P}[\mathbf{x}^{(n)}|\pi,\boldsymbol{\theta}] \right) - \log(T_n!) \right]. \tag{4}$$



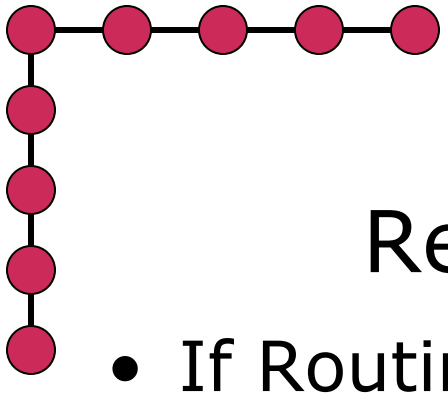
THE UNIVERSITY of WISCONSIN MADISON

Restricting the Solution Set

- If Routing is by shortest-path, not all permutations of co-occurrences result in data-feasible networks.
  - This is because for shortest path networks, all subpaths of a shortest path must also be shortest paths.



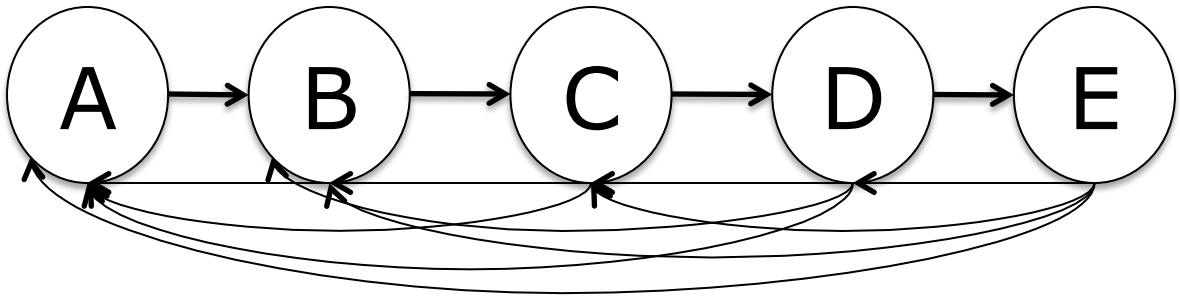
  - So if this is the shortest path from A to E, the following are also shortest paths: {ABCD}, {BCDE}, {ABC}, {BCD}, {CDE}, {AB}, {BC}, {CD}, {DE}.



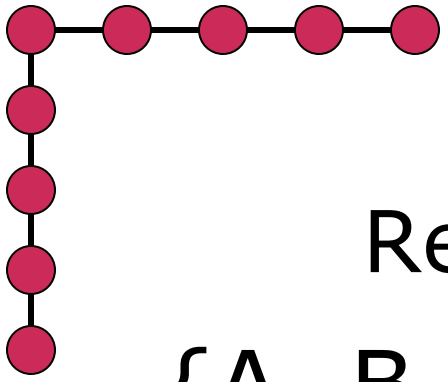
THE UNIVERSITY of WISCONSIN MADISON

Restricting the Solution Set

- If Routing is by shortest-path, not all permutations of co-occurrences result in data-feasible networks.
  - This is because for shortest path networks, all subpaths of a shortest path must also be shortest paths.



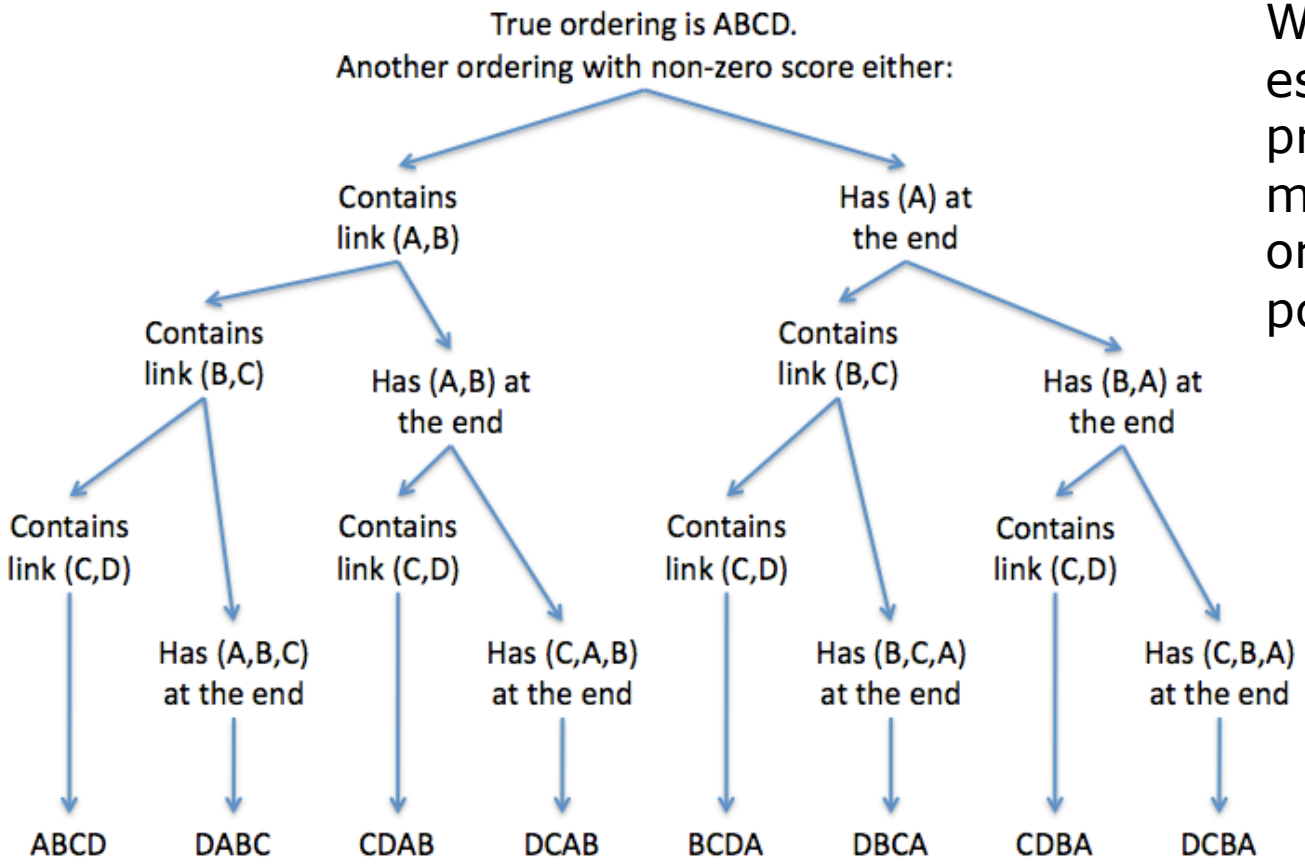
  - Therefore, when you have your resulting topology, there cannot be links from A->C, A->D, B->D, etc. All backwards links are still OK, but this still imposes a sizable restriction.



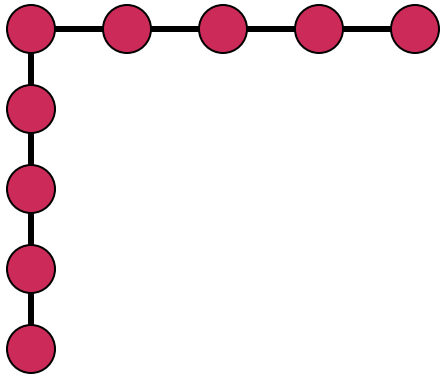
THE UNIVERSITY of WISCONSIN MADISON

Restricting the Solution Set

{A, B, C, D} ⇒ ? → ? → ? → ?



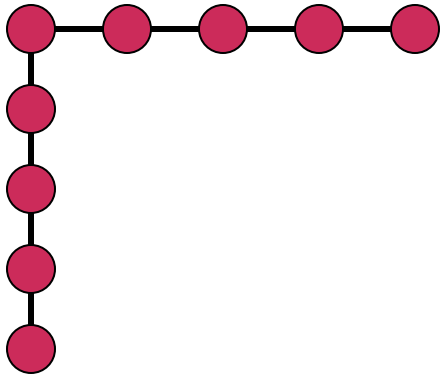
With a correct estimate of the probability transition matrix, only $2^{(n-1)}$ orderings will have positive likelihood.



THE UNIVERSITY of WISCONSIN MADISON

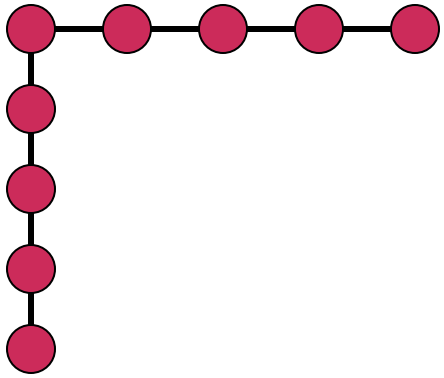
Future Questions

- How do we impose shortest path routing on the feasible set?
  - Is there a better way to think about the restriction when you don't have the correct signaling probabilities? Then all permutations may again have positive likelihood, but the resulting topology won't necessarily be feasible for shortest path.
- Can we extend the framework to tree-shaped pathways?
  - Network inference from co-occurrences is especially important in biological networks where signals do not take linear paths.



THE UNIVERSITY of WISCONSIN MADISON

Thank you!



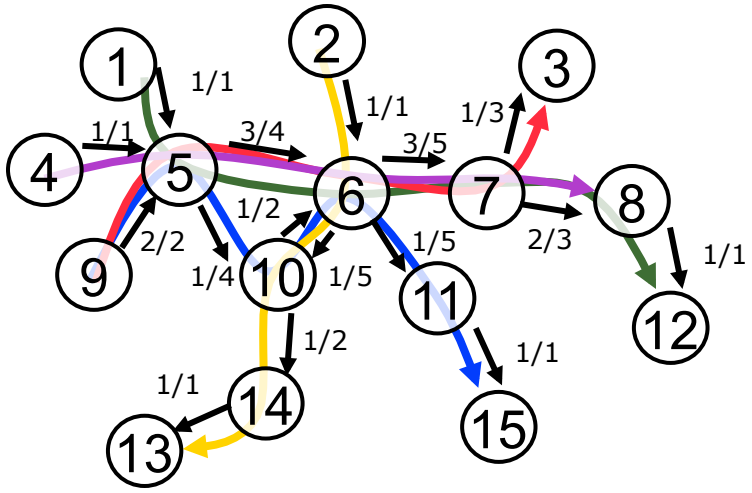
THE UNIVERSITY of WISCONSIN MADISON

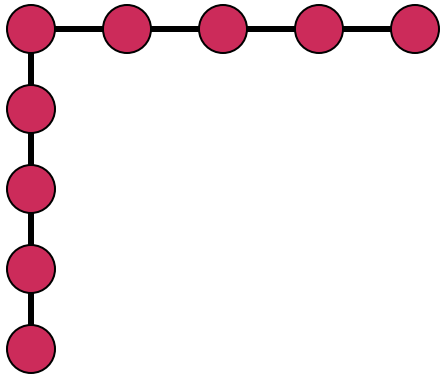
Extra Slides

Generative Model for Linear Pathways

- Model pathway generation by a random walk on the graph according to next-hop signaling probability matrix A:
  - The source node is drawn at random from the nodes in the graph.
  - The next hop is then drawn according to the transition probabilities.
  - This is the generative model– our assumption of how the co-occurrence observations are generated.

Obviously in a lot of practical or real-life scenarios, signaling does not occur as a random walk. But it turns out this is a useful way of modeling the problem!



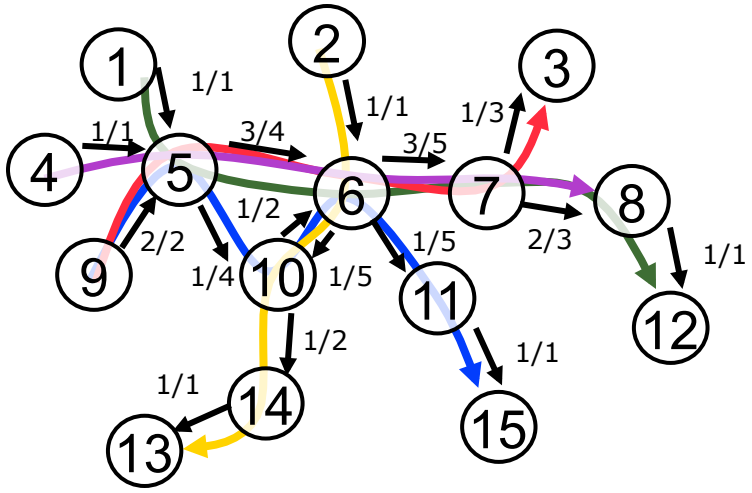


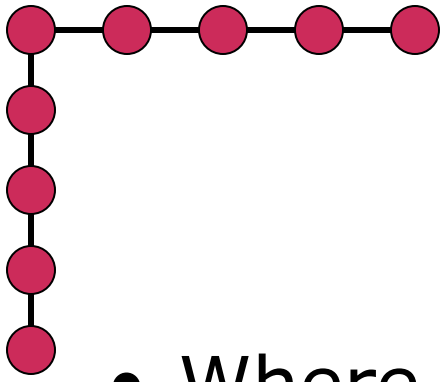
THE UNIVERSITY of WISCONSIN MADISON

Generative Model for Linear Pathways

- Model pathway generation by a random walk on the graph according to next-hop signaling probability matrix A.
- Then we can reduce the problem of finding topology and information flow to simply estimating the matrix A.
- Using this model, with knowledge of A, we can calculate the likelihood of any particular path given a co-occurrence observation.

For example if our co-occurrence is 2, 3, 6, 7, the order 2->6->7->3 has likelihood 1/5 and is the only order with nonzero likelihood.



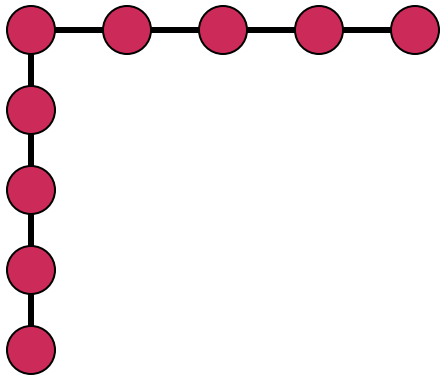


THE UNIVERSITY of WISCONSIN MADISON

Other Applications

- Where else is this result applicable?

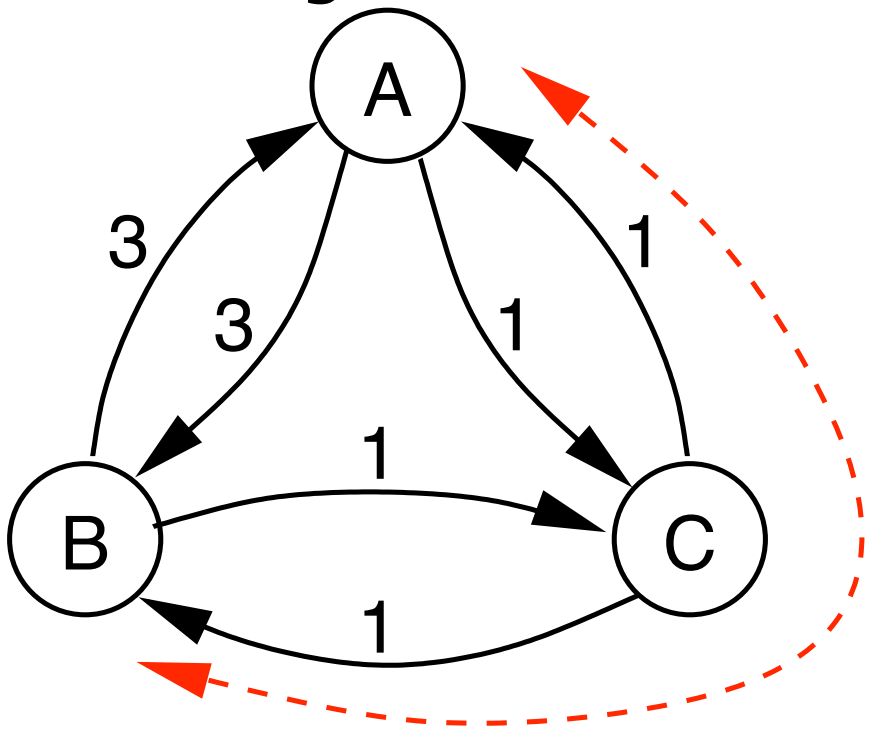
- Identifying misconfigured or misbehaving routers
- Exponential splitting– choosing link weights for OSPF for optimal load balancing
- Fast computation of shortest-paths in a network



THE UNIVERSITY of WISCONSIN MADISON

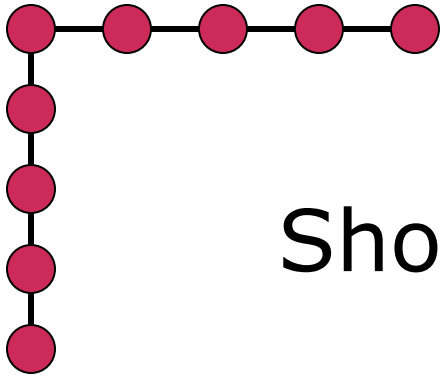
Shortest-Path Routing

- Let's look at P for a shortest-path routing network.



Routes

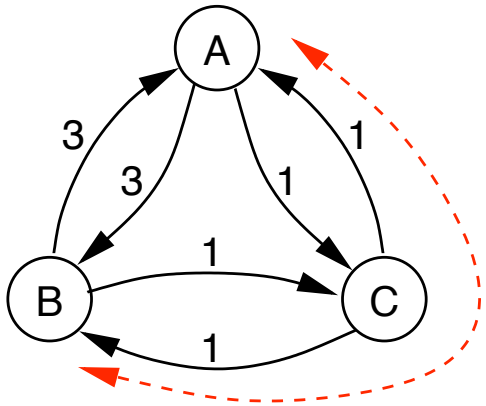
src	dst	route
A	B	A–C–B
A	C	A–C
B	A	B–C–A
B	C	B–C
C	A	C–A
C	B	C–B



THE UNIVERSITY of WISCONSIN MADISON

Shortest-Path Routing Example

- Let's look at P for a shortest-path routing network.

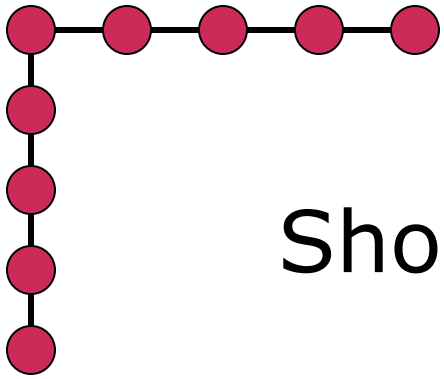


Routes

src	dst	route
A	B	A–C–B
A	C	A–C
B	A	B–C–A
B	C	B–C
C	A	C–A
C	B	C–B

$$D = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix}$$

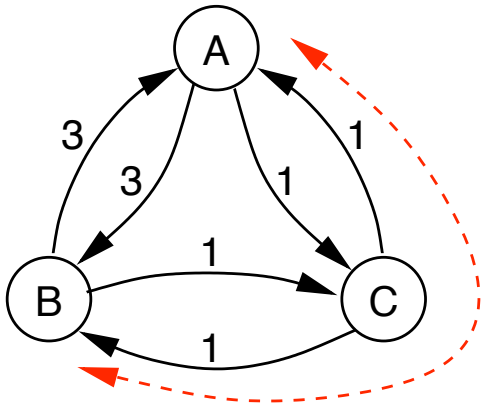
$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$



THE UNIVERSITY of WISCONSIN MADISON

Shortest-Path Routing Example

- Now say we get the co-occurrence "A, B, C". What is the score of every ordering?

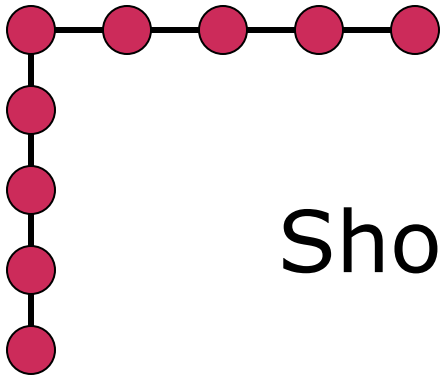


Routes

src	dst	route
A	B	A–C–B
A	C	A–C
B	A	B–C–A
B	C	B–C
C	A	C–A
C	B	C–B

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

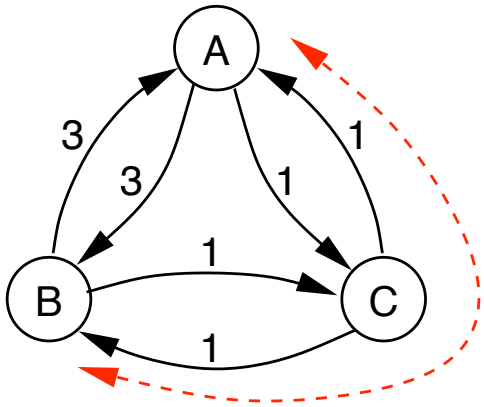
ordering	Example 1
A–B–C	$0 \times 1 = 0$
A–C–B	$1 \times 1/2 = 1/2$
B–A–C	$0 \times 1 = 0$
B–C–A	$1 \times 1/2 = 1/2$
C–A–B	$1/2 \times 0 = 0$
C–B–A	$12/ \times 0 = 0$



THE UNIVERSITY of WISCONSIN MADISON

Shortest-Path Routing Example

- Now say we get the co-occurrence "A, B, C"

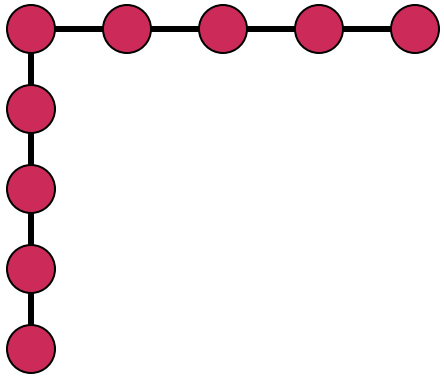


Routes

src	dst	route
A	B	A–C–B
A	C	A–C
B	A	B–C–A
B	C	B–C
C	A	C–A
C	B	C–B

ordering	Example 1
A–B–C	$0 \times 1 = 0$
A–C–B	$1 \times 1/2 = 1/2$
B–A–C	$0 \times 1 = 0$
B–C–A	$1 \times 1/2 = 1/2$
C–A–B	$1/2 \times 0 = 0$
C–B–A	$12/ \times 0 = 0$

- If we have information about the first node, then the *only* permutation with positive score is the correct one!

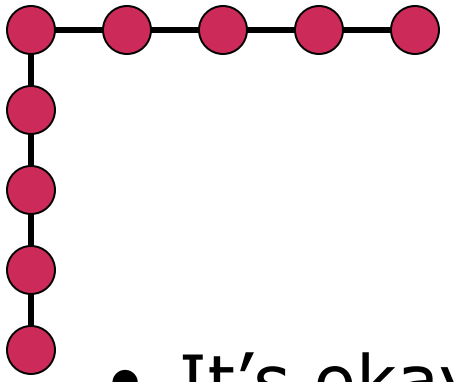


THE UNIVERSITY of WISCONSIN MADISON

Very General

- This result that the only permutation with positive score actually holds under very general conditions.

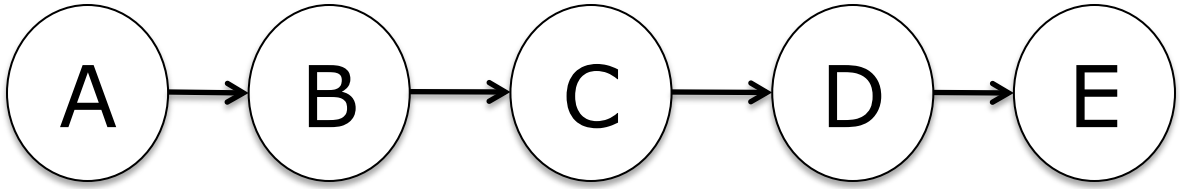
- It’s okay if the network has asymmetric links.
- It’s okay if the routing isn’t shortest-path, as long as it follows a “nested” routing policy.
- It’s okay if there are equal-cost paths, as long as the cost metric is hop-count.



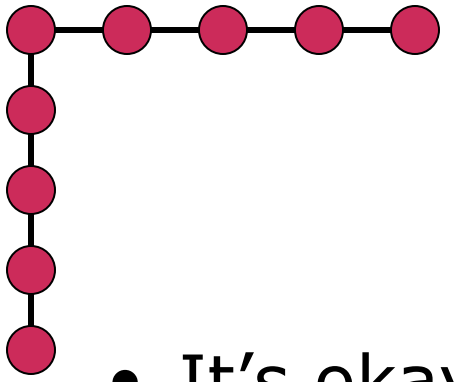
THE UNIVERSITY of WISCONSIN MADISON

Nested Routing

- It's okay if the routing isn't shortest-path, as long as it follows a "nested" routing policy.
  - When we think of shortest-path routing, we know that shortest paths are made up of shortest paths:



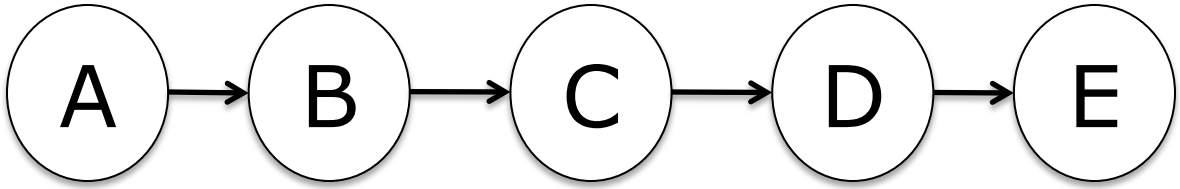
  - So if this is the shortest path from A to E, the following are also shortest paths: {ABCD}, {BCDE}, {ABC}, {BCD}, {CDE}, {AB}, {BC}, {CD}, {DE}.



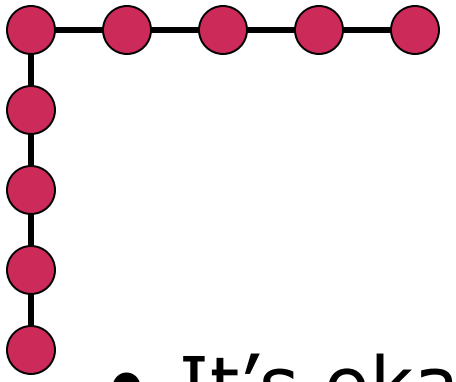
THE UNIVERSITY of WISCONSIN MADISON

Nested Routing

- It's okay if the routing isn't shortest-path, as long as it follows a "nested" routing policy.
  - When we think of shortest-path routing, we know that shortest paths are made up of shortest paths:



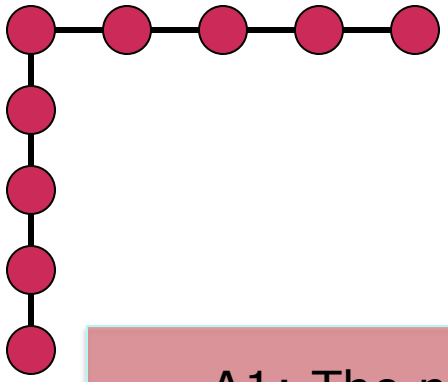
  - As long as this principle holds– that paths chosen by the policy are made up of other chosen paths– then the results hold.



THE UNIVERSITY of WISCONSIN MADISON

Equal-Cost Paths

- It's okay if there are equal-cost paths, as long as the cost metric is hop-count.
  - Hop count is often the cost metric. For example, the version of RIP (Routing Information Protocol) implemented in Cisco routers uses hop count by default.
  - Even if hop count is not the cost metric, the result actually still holds if you use the transition matrix of the dual graph, or a link-to-link transition matrix.

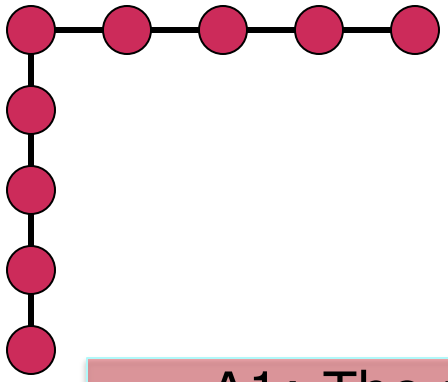


THE UNIVERSITY of WISCONSIN MADISON

Theorem 1

- A1: The network uses a nested routing policy.
- A2: $X = \{x_1 , \dots , x_n\}$ is an unordered co-occurrence observation which consists of nodes from a true route in the network.
- A3: The first node in the route is known; we denote this by x_s
- A4: If there are equal-cost paths, the cost metric is hop count.

- Then an ordering $\pi = [s, \dots]$ is the correct ordering for the route if and only if the score, derived from the transition matrix P, of the route $\{x_s, x_{\pi(2)} , \dots , x_{\pi(n)}\}$ is positive.

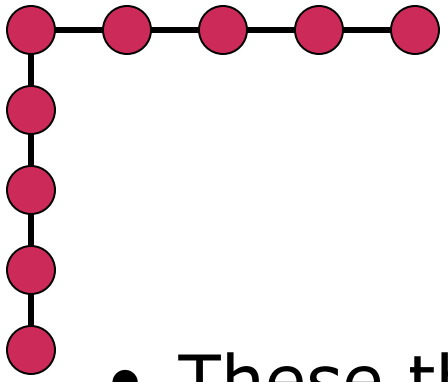


THE UNIVERSITY of WISCONSIN MADISON

Theorem 2

- A1: The network uses a nested routing policy.
- A2: $X = \{x_1 , \dots , x_n\}$ is an unordered co-occurrence observation which consists of nodes from a true route in the network.
- A3: The first node in the route is known; we denote this by x_s

- Then an ordering $\pi = [s, \dots]$ is the correct ordering for the route if and only if the score, derived from the transition matrix Q built from link-to-link transitions, of the route $\{x_s, x_{\pi(2)} , \dots , x_{\pi(n)}\}$ is positive.

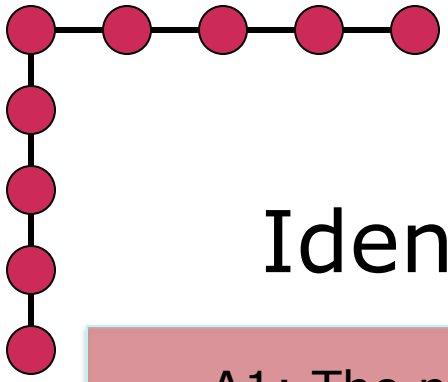


THE UNIVERSITY of WISCONSIN MADISON

Other Applications

- These theorems solve the co-occurrence ordering problem for graphs with nested routing policies. Where else is this result applicable?

- Identifying misconfigured or misbehaving routers
- Exponential splitting– choosing link weights for OSPF for optimal load balancing
- Fast computation of shortest-paths in a network

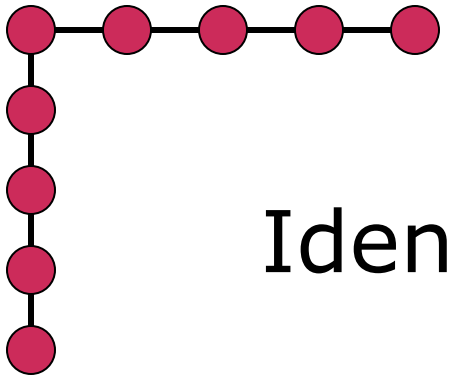


THE UNIVERSITY of WISCONSIN MADISON

Identifying Misbehaving Routers

- A1: The network uses a nested routing policy.
- A2: $X = \{x_1 , \dots , x_n\}$ is an unordered co-occurrence observation which consists of nodes from a true route in the network.
- A3: The first node in the route is known; we denote this by x_s
- A4: If there are equal-cost paths, the cost metric is hop count.

- Then an ordering $\pi = [s, \dots]$ is the correct ordering for the route if and only if the score, derived from the transition matrix P, of the route $\{x_s, x_{\pi(2)} , \dots , x_{\pi(n)}\}$ is positive.

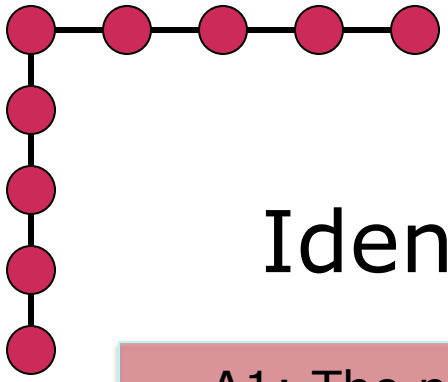


THE UNIVERSITY of WISCONSIN MADISON

Identifying Misbehaving Routers

- Then an ordering π = [s, …] is the correct ordering for the route if and only if the score, derived from the transition matrix P, of the route $\{x_s, x_{\pi(2)}, \ldots, x_{\pi(n)}\}$ is positive.

- It seems that if a router reports a route with ordering π which is incorrect, we can identify it immediately by looking at the score of that route...

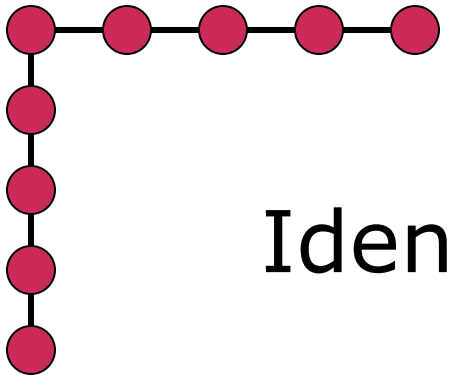


THE UNIVERSITY of WISCONSIN MADISON

Identifying Misbehaving Routers

- A1: The network uses a nested routing policy.
- A2: $X = \{x_1 , \ldots , x_n\}$ is an unordered co-occurrence observation which consists of nodes from a true route in the network.
- A3: The first node in the route is known; we denote this by x_s
- A4: If there are equal-cost paths, the cost metric is hop count.

- … but the second assumption requires that the co-occurrence observation consists of nodes from a true route in the network.
- If a router misbehaves by sending a route that has nodes that aren't in a real route, then the theorems don't hold.



THE UNIVERSITY of WISCONSIN MADISON

Identifying Misbehaving Routers

- If a router misbehaves by sending a route that has nodes that aren't in a real route, then the theorems don't hold.

- The next step is to look more carefully at this case and see how likely it is for a router to report a route that cannot be identified as problematic.
- My hunch is that the router has to work pretty hard to fool us.