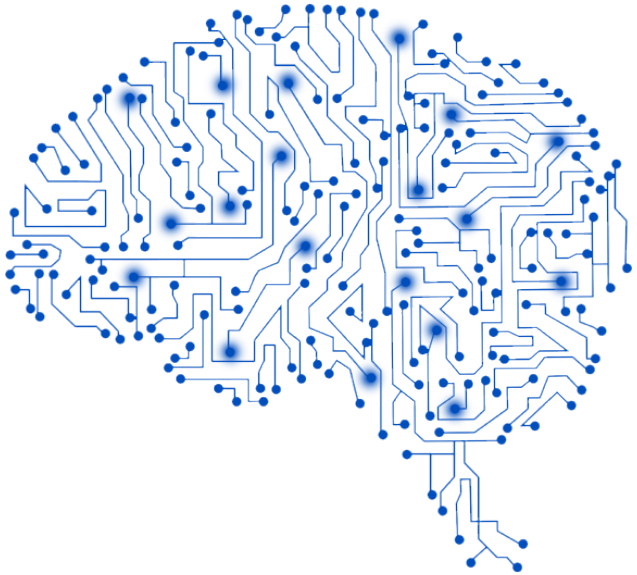
# Master's Thesis

# Motif Discovery for Connectomics: Finding Computational Units in the Brain's Wiring Diagram

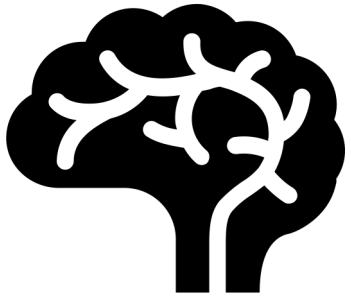


June 26th, 2019

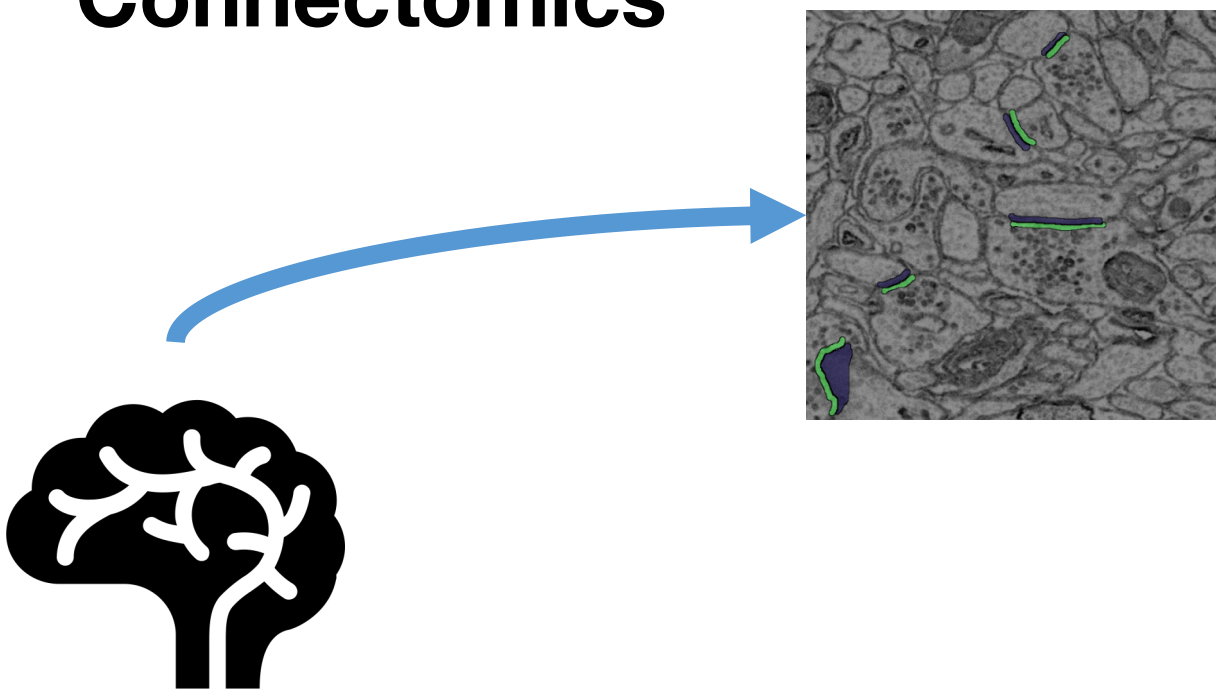
Presenter : Alleon Antoine, CSE – Master 2

École Polytechnique Fédérale de Lausanne – EPFL
Visual Computing Group, Harvard SEAS

# Connectomics

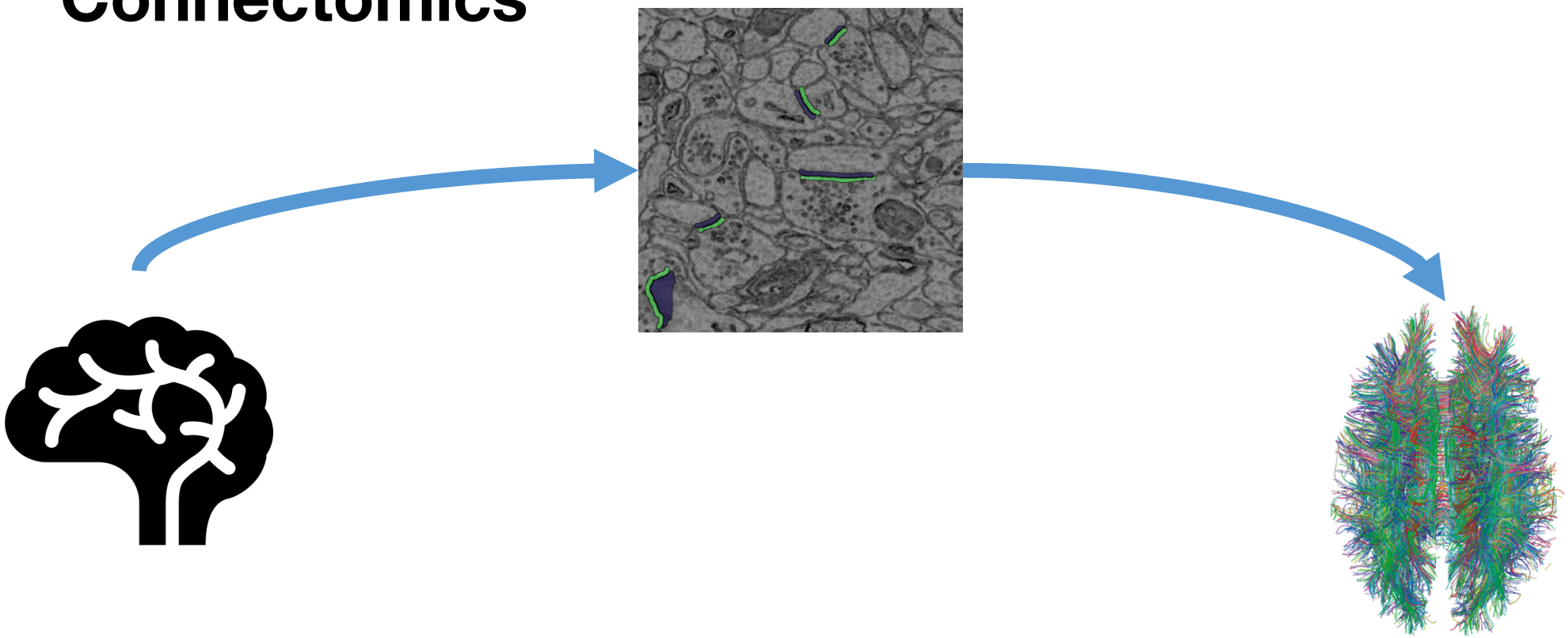


# Connectomics

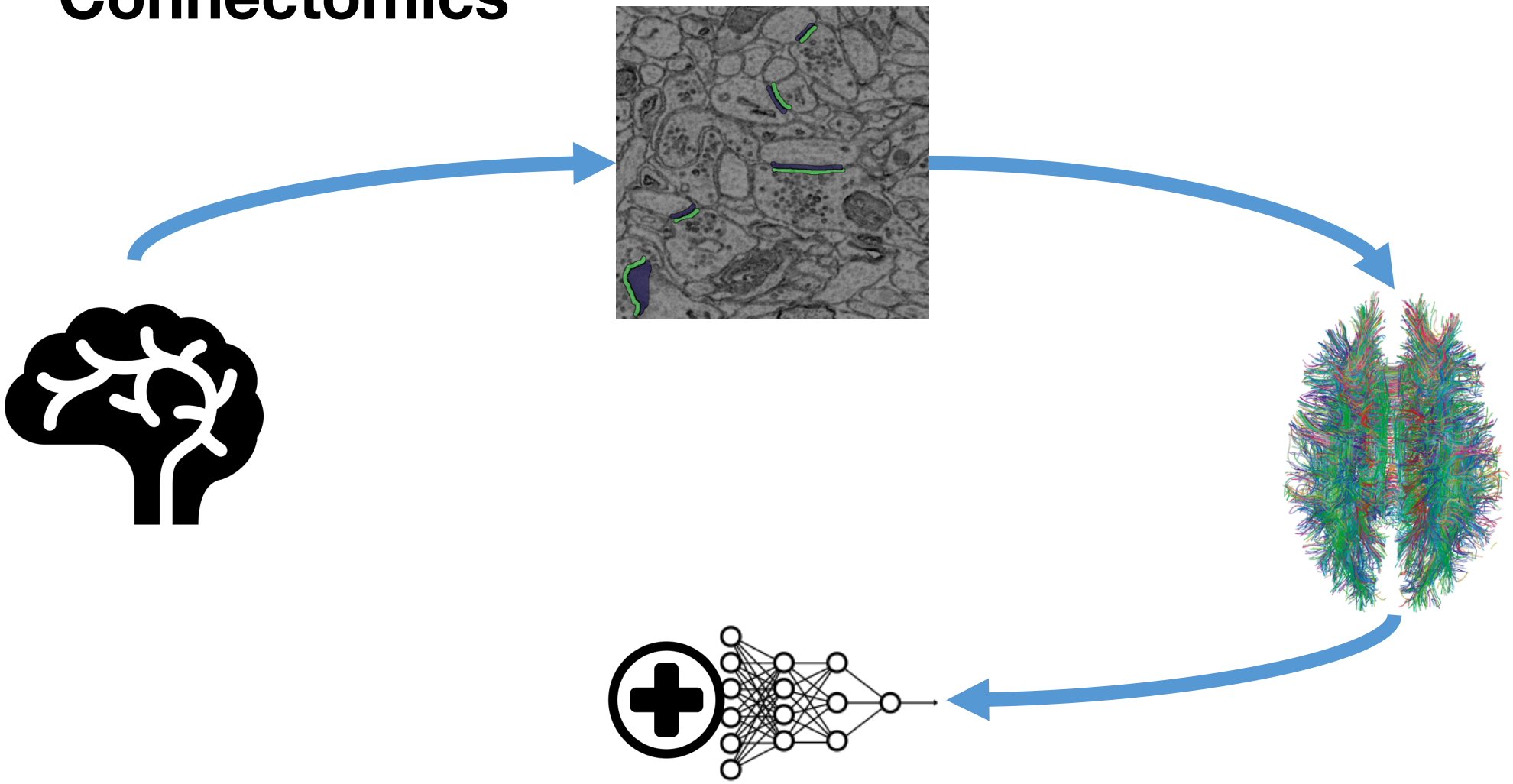


D. Haehn et al. Scalable interactive visualization for connectomics. *Informatics.*

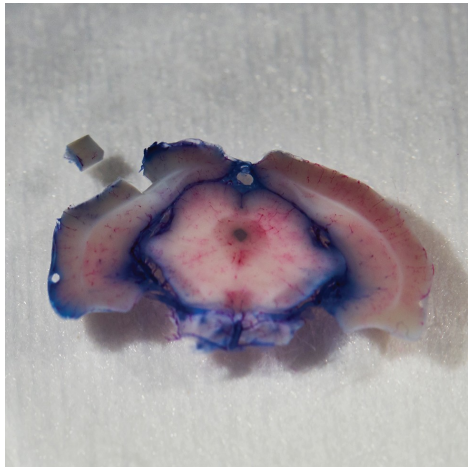
# Connectomics



# Connectomics



# Connectomics

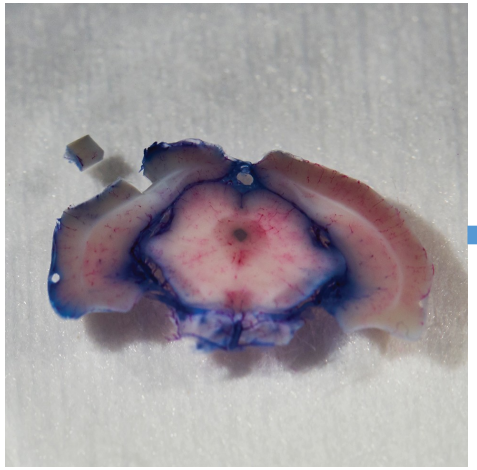


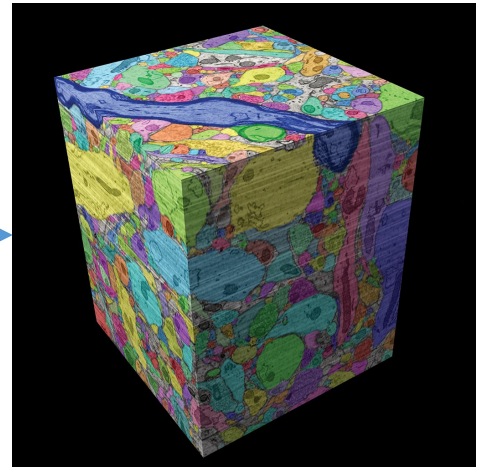
Bio-engineering:

- Brain cutting
- Brain imaging

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

# Connectomics





Bio-engineering:

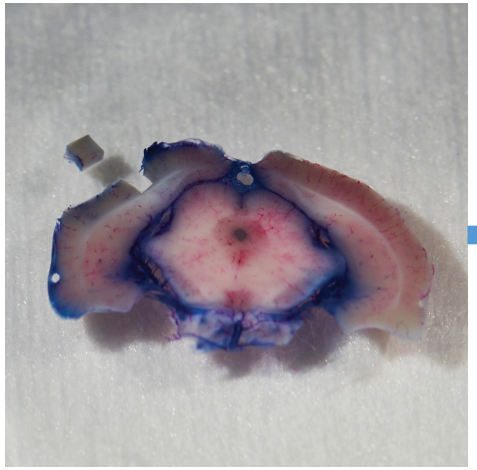
- Brain cutting
- Brain imaging

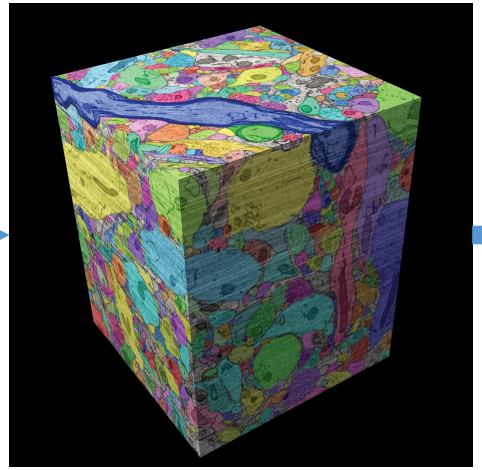
Computer Sciences:

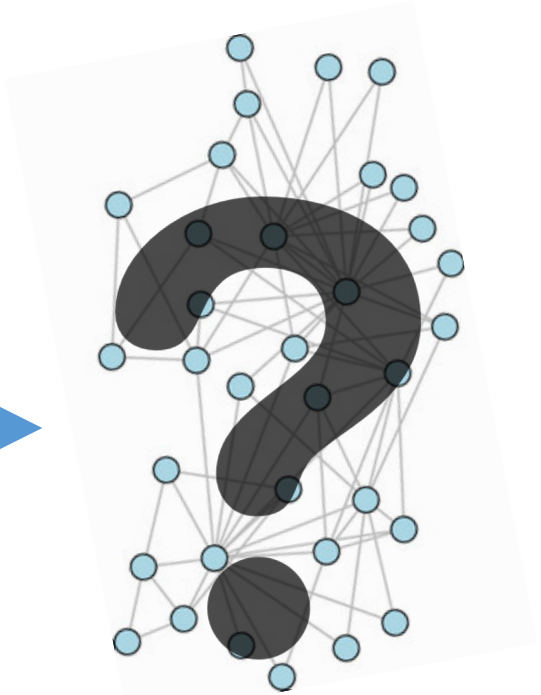
- Alignment of images
- Segmentation

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

# Connectomics







Bio-engineering:

- Brain cutting
- Brain imaging

Computer Sciences:

- Alignment of images
- Segmentation

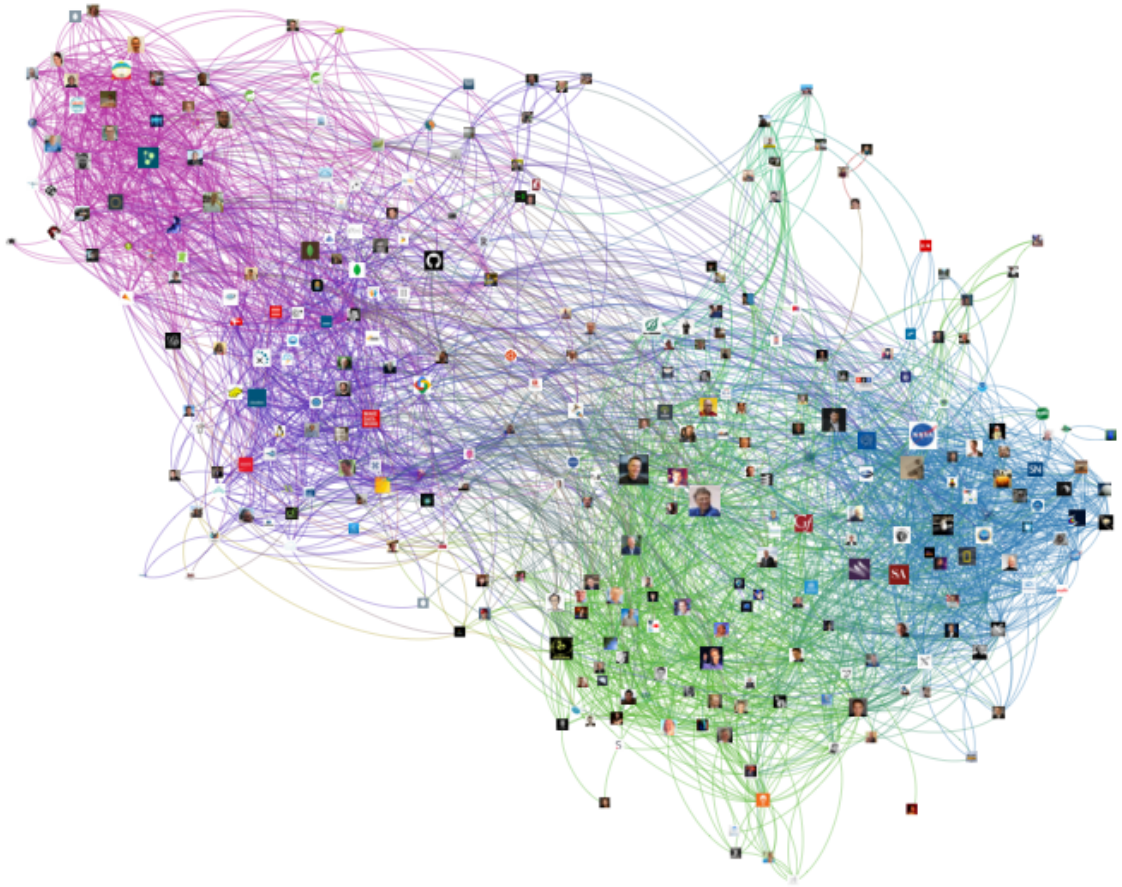
Applications:

- Graph analysis
- Visualization

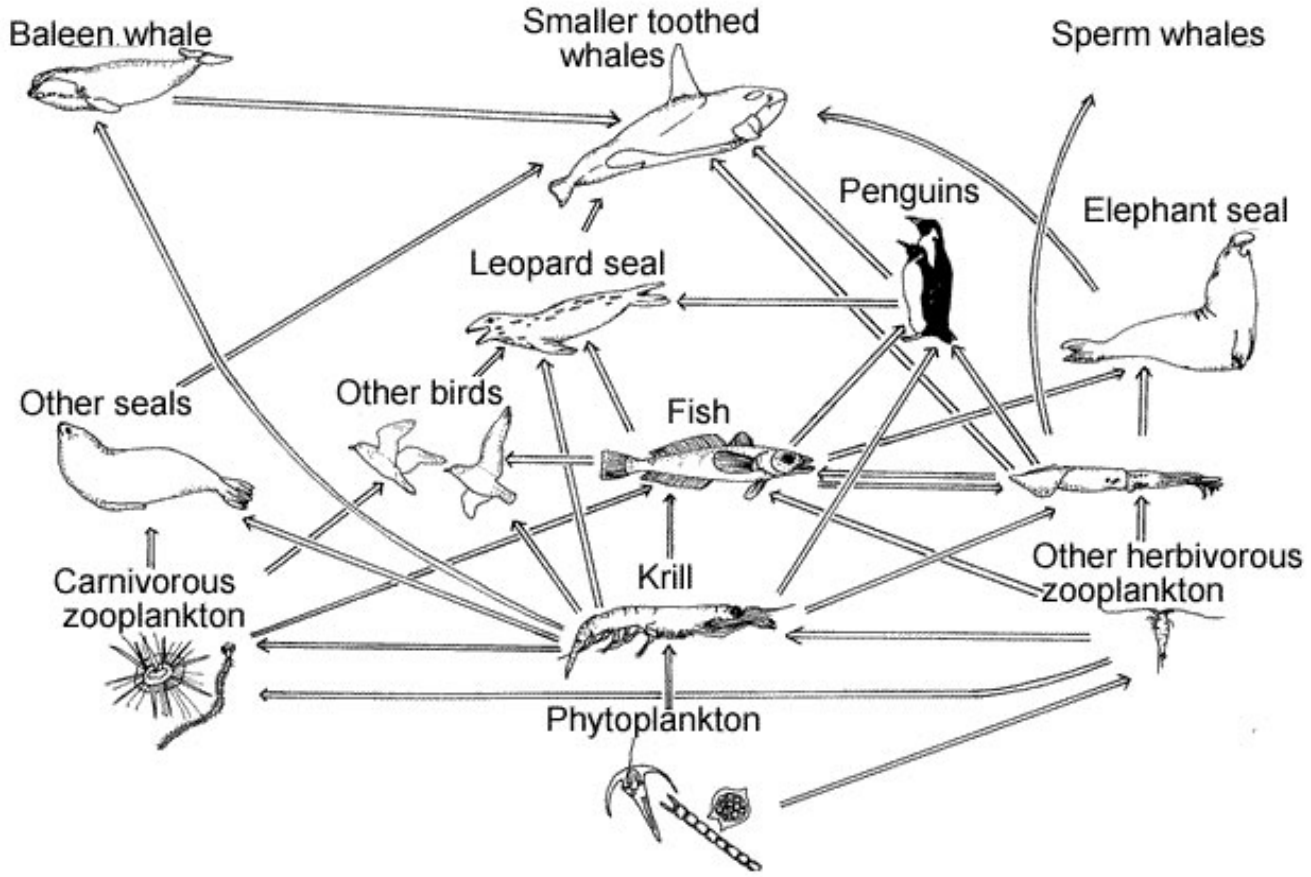
Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

# Motif Discovery

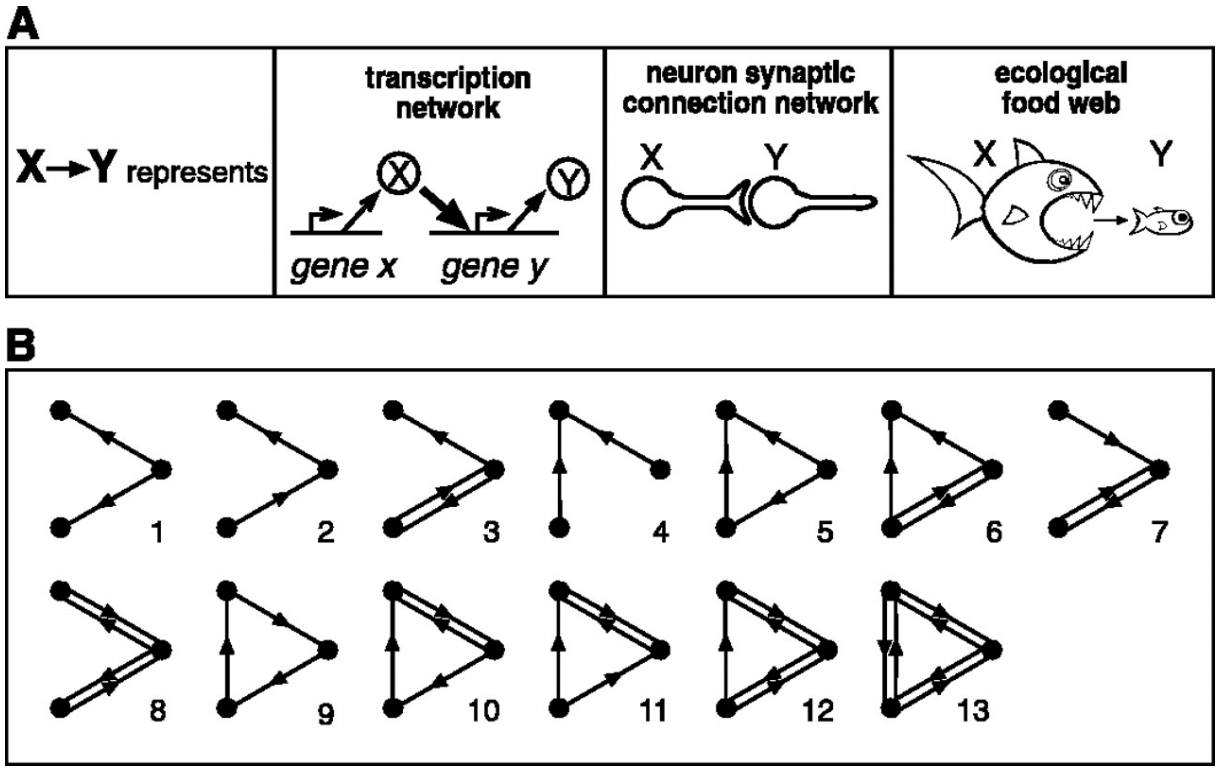
# Motif Discovery



# Motif Discovery



# Motif Discovery




Gene regulation

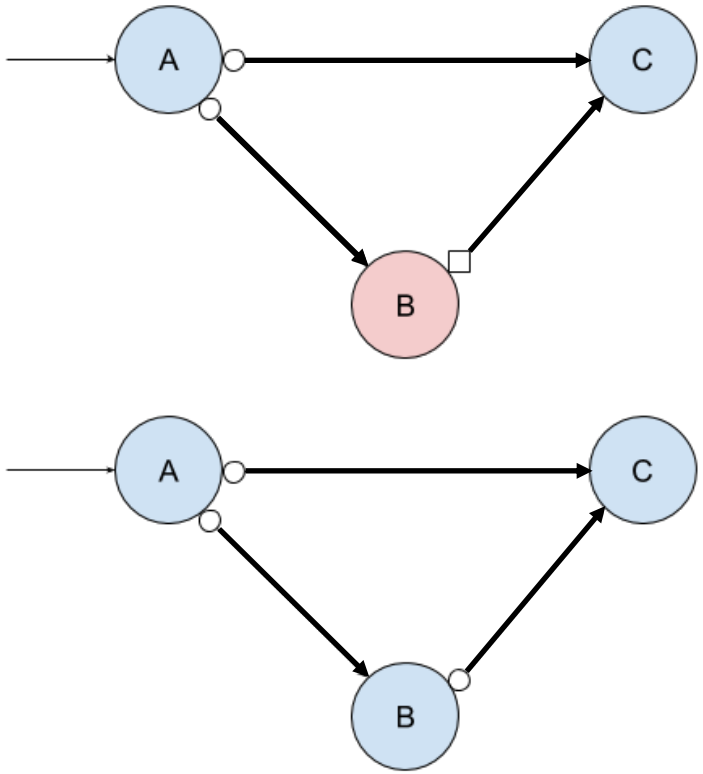
Metabolic system

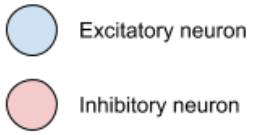
Voter networks

Neuronal connections

R. Milo et al.: Network motifs: simple building blocks of complex networks. *Science*

# Motif Discovery

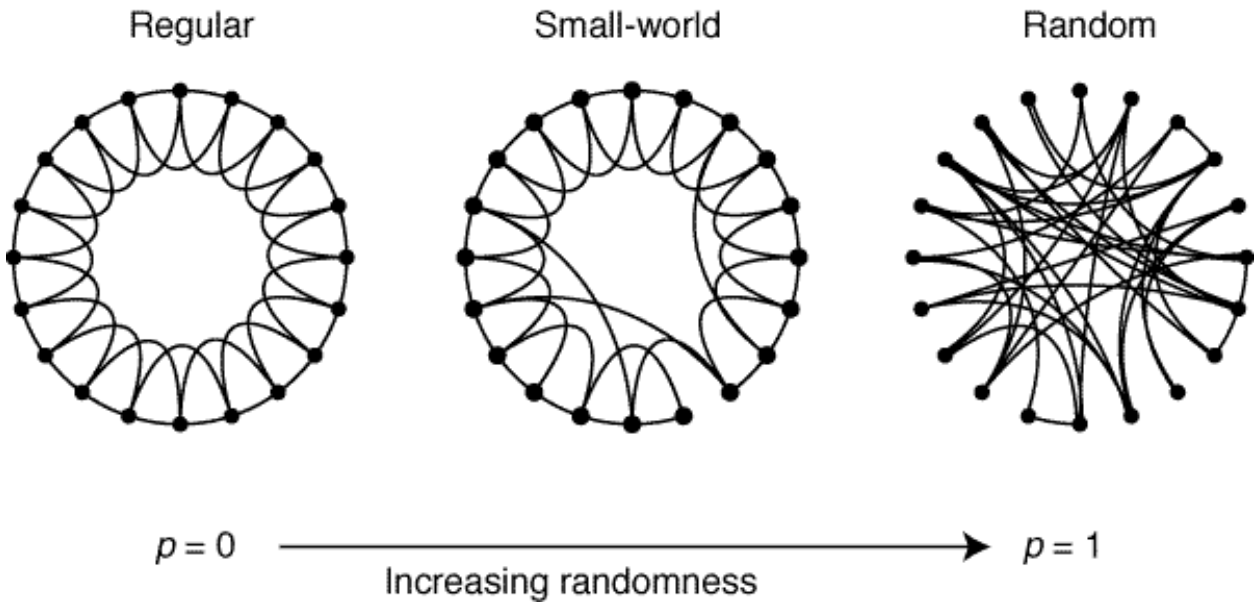



Feed-Forward loop motif

Two types of neurons

# Motif Discovery




Small-world properties of neural networks

Highly clustered

Small path length

B. Uzzi et al.: Small-world networks and management science research: A review.

# Agenda

**Survey & Application**: Small motifs (3 to 5 nodes)

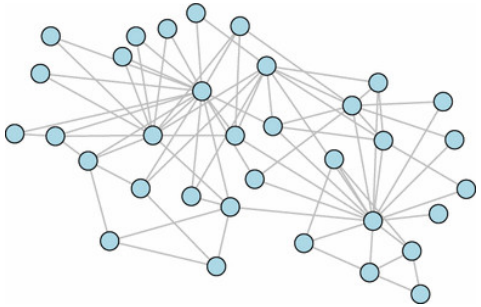
- Brain wiring diagram
- Computational performance

**Exploration**: Large motifs (6 and more nodes)

- Improve efficiency on large motif detection

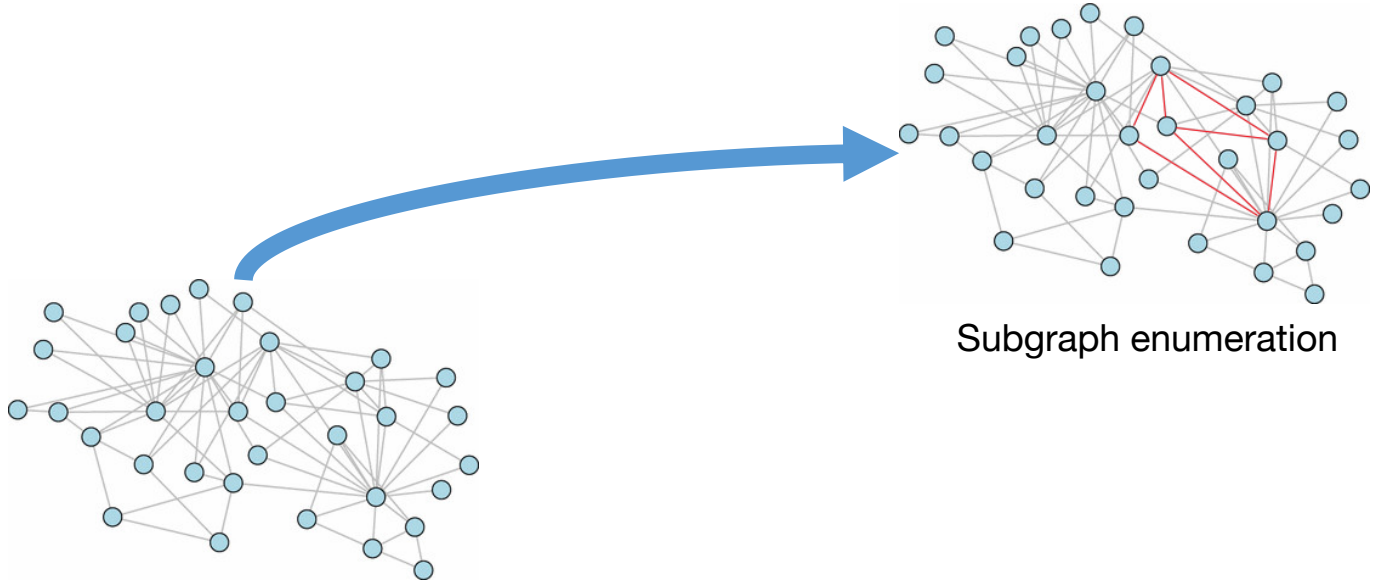
# Small Subgraph enumeration: Survey and applications

# Motif discovery - Intuition

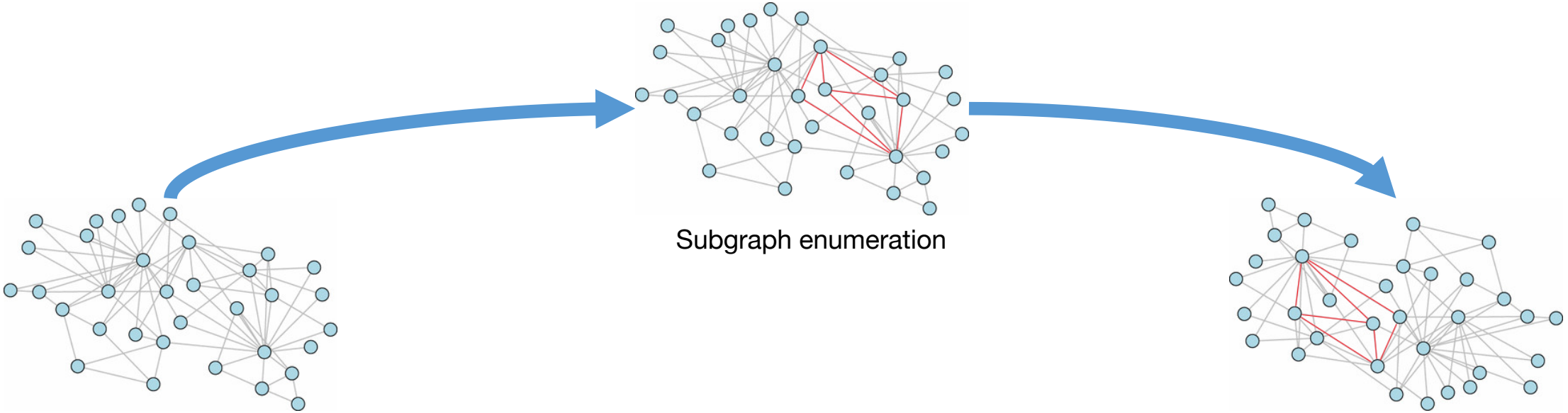


Initial graph

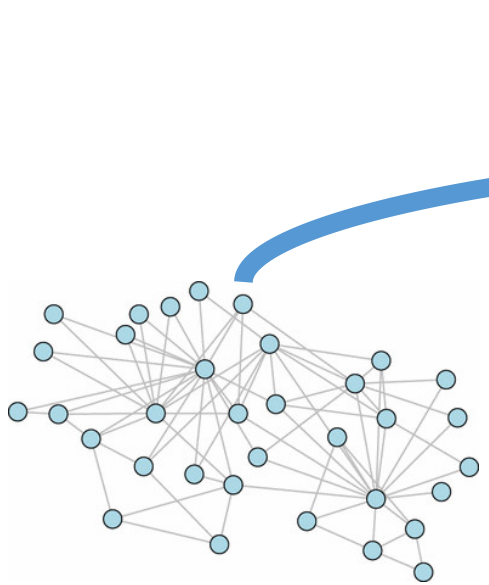
# Motif discovery - Intuition



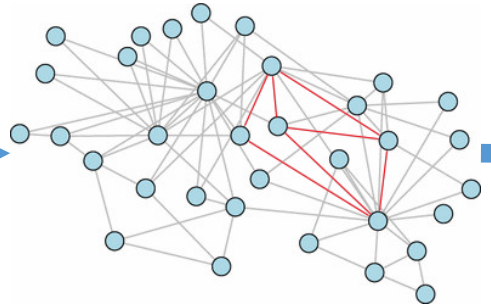
# Motif discovery - Intuition



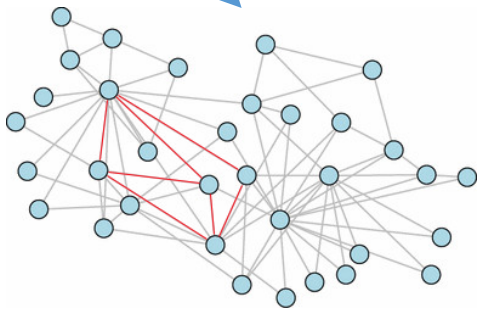
# Motif discovery - Intuition



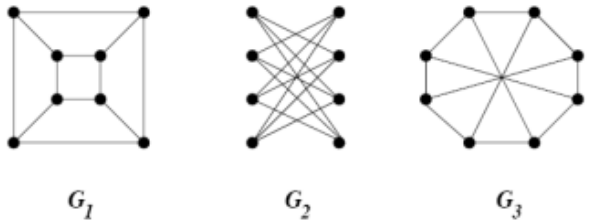
Initial graph



Subgraph enumeration

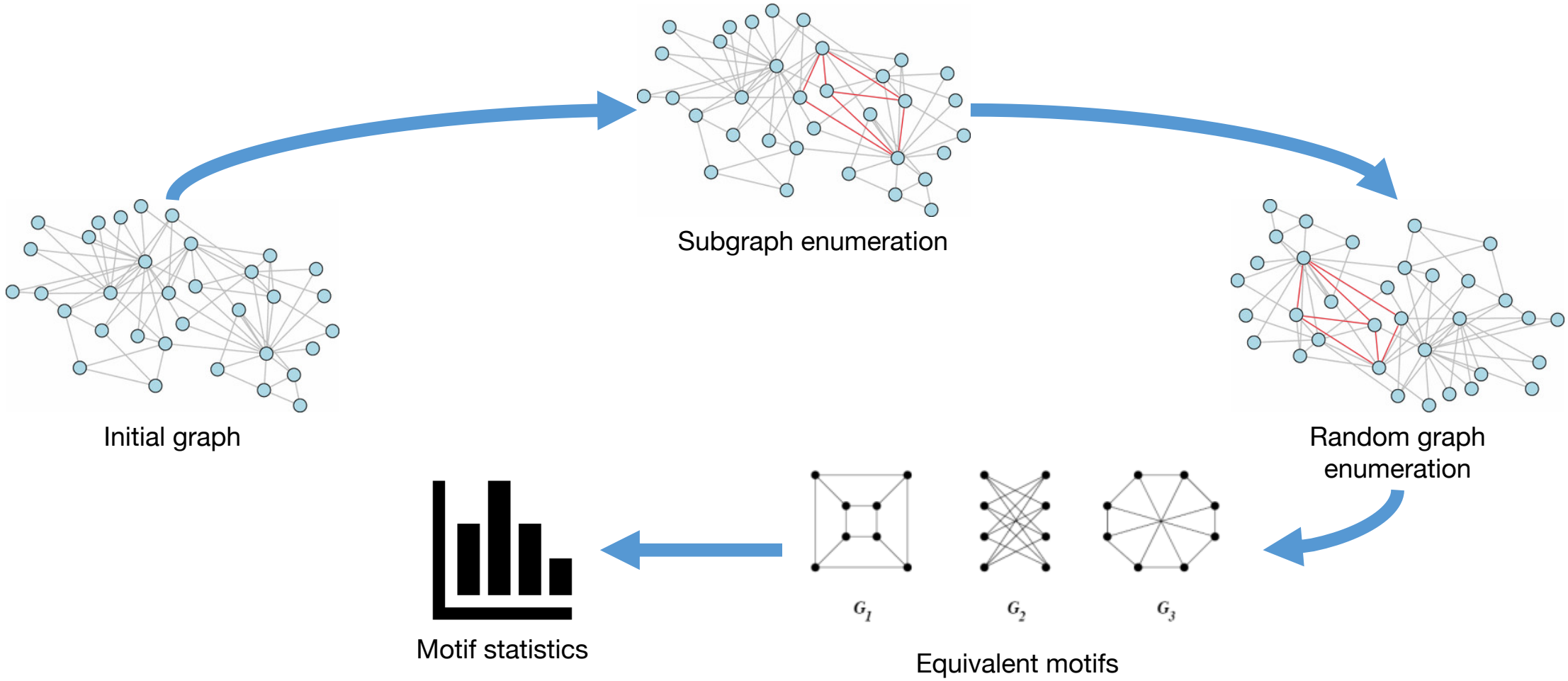


Random graph enumeration




Equivalent motifs

# Motif discovery - Intuition



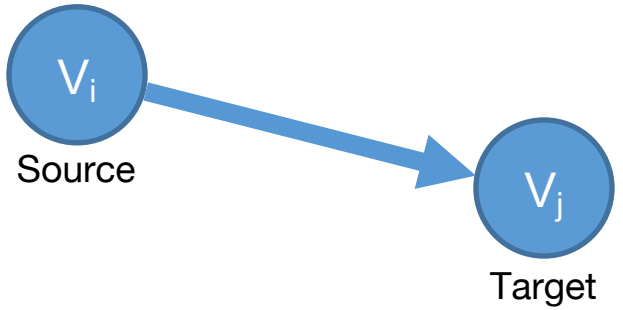
# Methods – Graph Notation

Graph - $G = \{V, E\}$

# Methods – Graph Notation

Graph - $G = \{V, E\}$

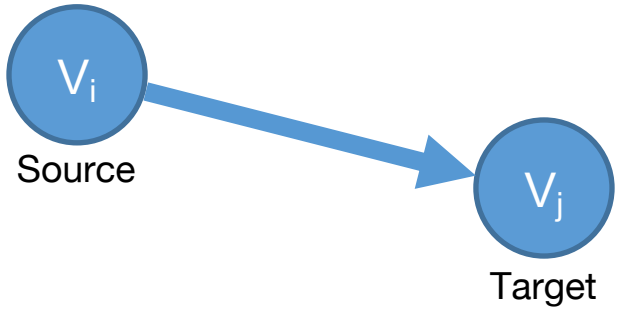
Edge - $E_{ij} = (V_i, V_j)$ with $V_i, V_j \in V, i \neq j$



# Methods – Graph Notation

Graph - $G = \{V, E\}$

Edge - $E_{ij} = (V_i, V_j)$ with $\mathrm{V_i}, \mathrm{V_j} \in V, i \neq j$

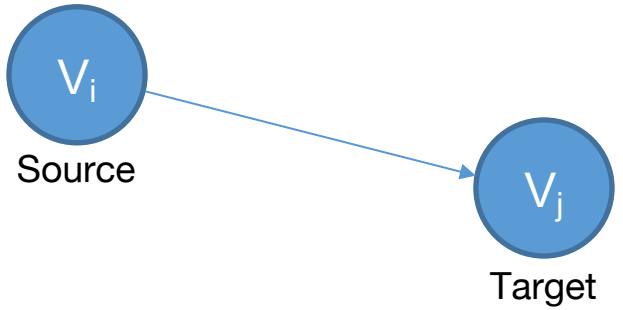



Subgraph - $G' = \{V', E'\}$ with $V' \subseteq V; \mathrm{E}' \subseteq (V', V') \cap E$

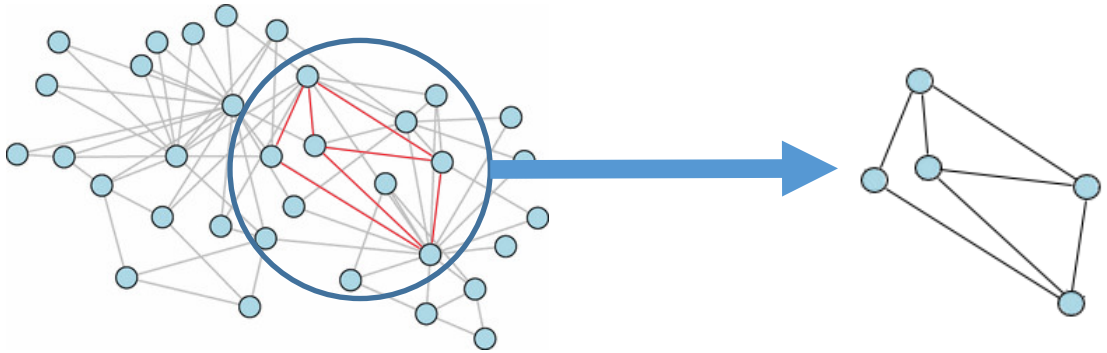
# Methods – Graph Notation

Graph - $G = \{V, E\}$

Edge - $E_{ij} = (V_i, V_j)$ with $\mathrm{V_i}, \mathrm{V_j} \in V, i \neq j$

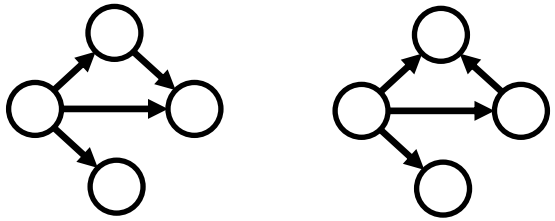



Subgraph - $G' = \{V', E'\}$ with $V' \subseteq V; \mathrm{E}' \subseteq (V', V') \cap E$



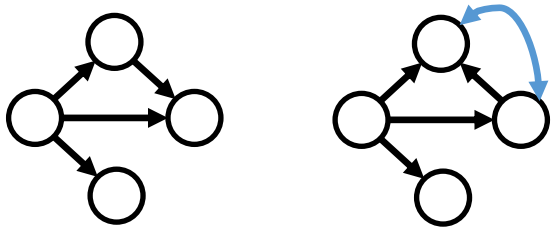
# Methods – Graph Notation

Isomorphic group - $G_1 \approx G_2$: $\exists F;\ F(V_1) = V_2,\ E_2 = \tilde{F}(E_1) = (F(V_{1_i}), F(V_{1_j}))$



# Methods – Graph Notation

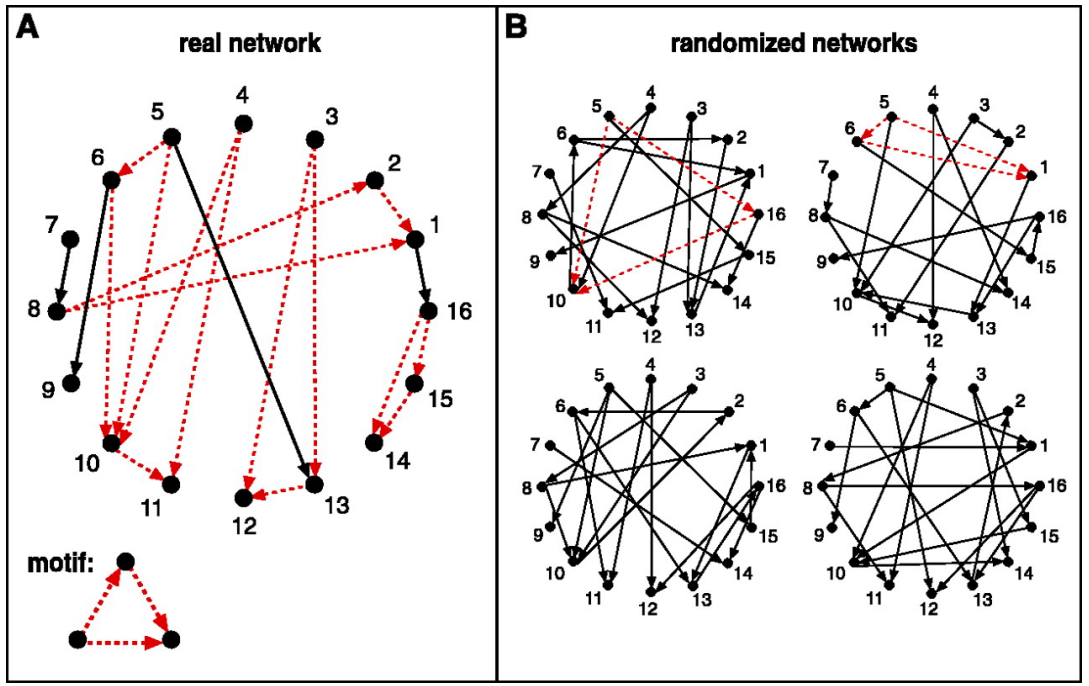
Isomorphic group - $G_1 \approx G_2$: $\exists F;\ F(V_1) = V_2,\ E_2 = \tilde{F}(E_1) = (F(V_{1_i}), F(V_{1_j}))$



# Methods – Graph Notation

Isomorphic group - $G_1 \approx G_2 : \exists F;\ F(V_1) = V_2,\ E_2 = \tilde{F}(E_1) = (F(V_{1_i}), F(V_{1_j}))$

Z-Score - $Z_m = (f_m - \mu_m)/\sigma_m$

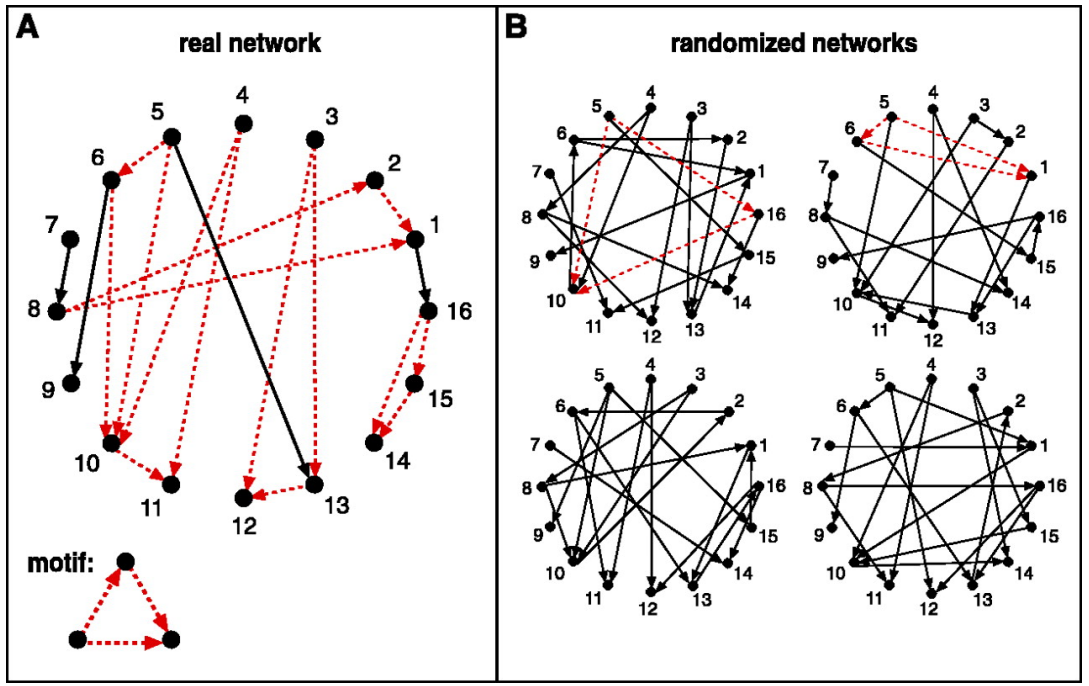


# Methods – Graph Notation

Isomorphic group - $G_1 \approx G_2 : \exists F;\ F(V_1) = V_2,\ E_2 = \tilde{F}(E_1) = (F(V_{1_i}), F(V_{1_j}))$

Z-Score - $Z_m = (f_m - \mu_m)/\sigma_m$

$F_o = 5$
$F_1 = 1$
$F_2 = 1$
$F_3 = 0$
$F_4 = 0$

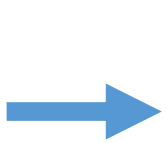


# Methods – Graph Notation

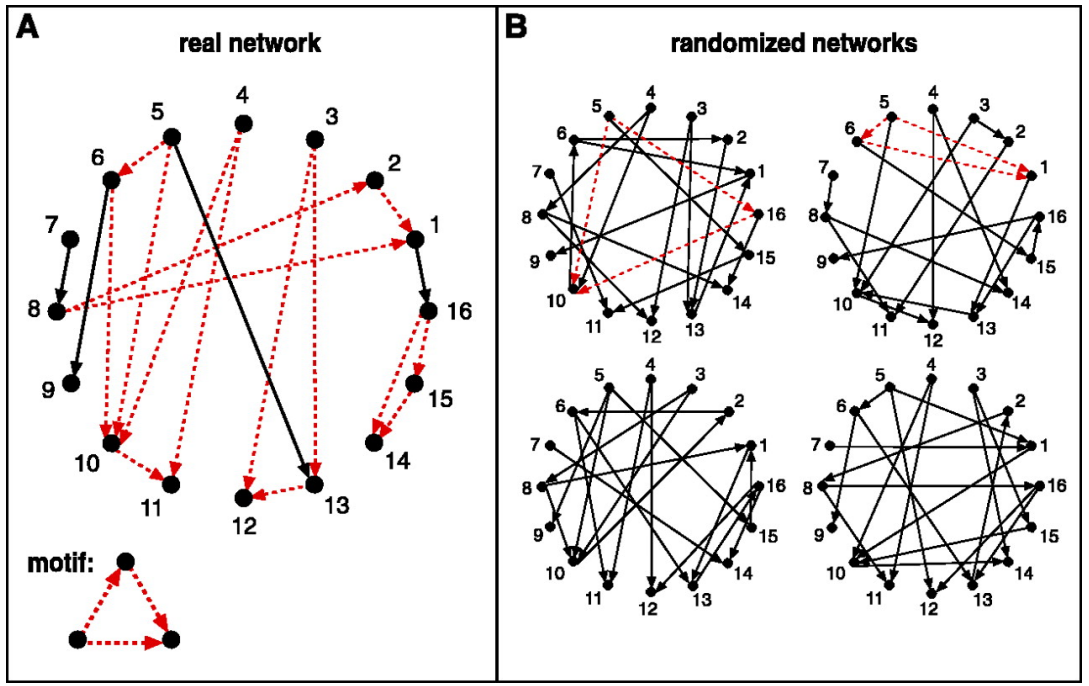
Isomorphic group - $G_1 \approx G_2 : \exists F;\ F(V_1) = V_2,\ E_2 = \tilde{F}(E_1) = (F(V_{1_i}), F(V_{1_j}))$

Z-Score - $Z_m = (f_m - \mu_m)/\sigma_m$

$F_o = 5$
$F_1 = 1$
$F_2 = 1$
$F_3 = 0$
$F_4 = 0$



$f = 5$
$\mu = 0.5$
$\sigma = 0.5$



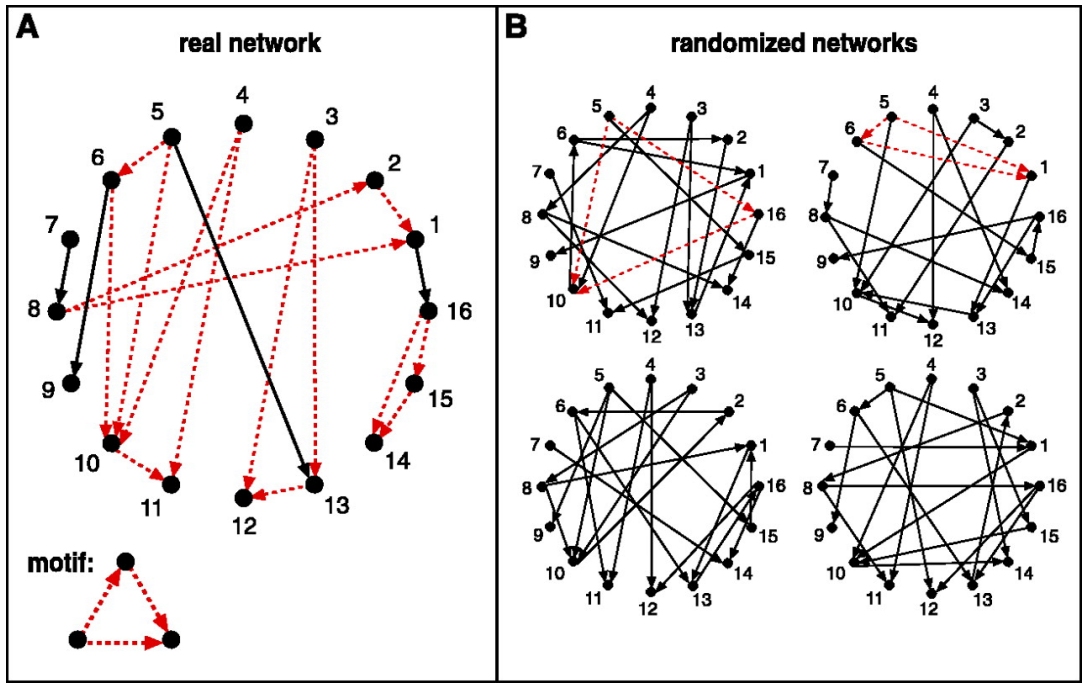
# Methods – Graph Notation

Isomorphic group - $G_1 \approx G_2 : \exists F;\ F(V_1) = V_2,\ E_2 = \tilde{F}(E_1) = (F(V_{1_i}), F(V_{1_j}))$

Z-Score - $Z_m = (f_m - \mu_m)/\sigma_m$

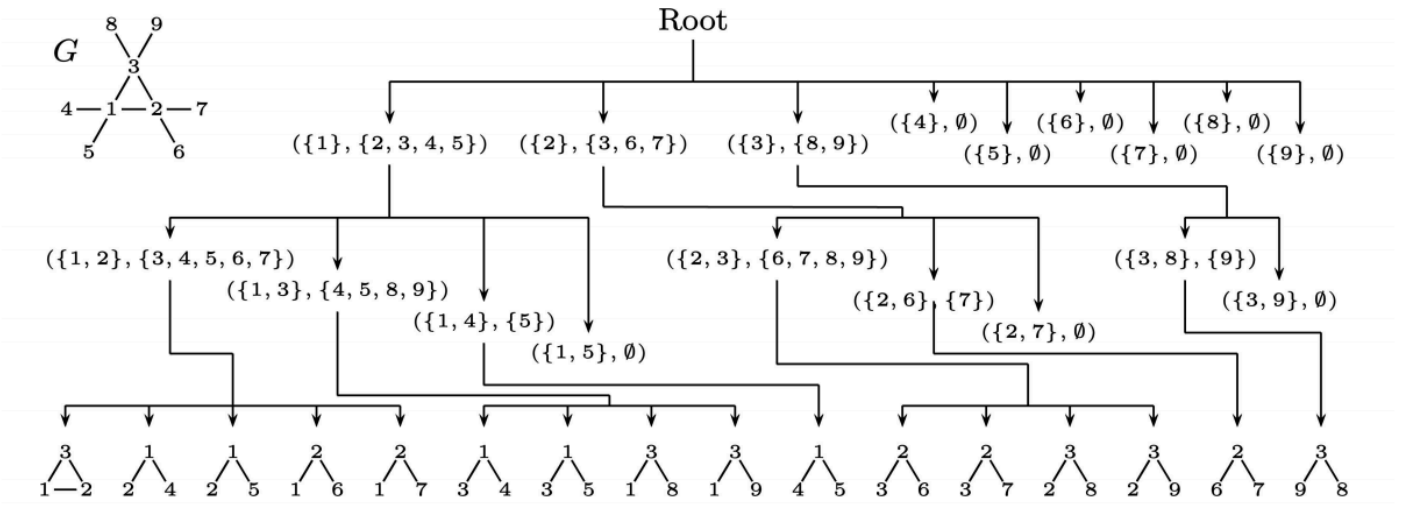
$F_o = 5$
$F_1 = 1$
$F_2 = 1$
$F_3 = 0$
$F_4 = 0$

→ $f = 5$, $\mu = 0.5$, $\sigma = 0.5$ → $Z = 9$



# Motif discovery - History

- *Nauty* – 1981
- *Mfinder* – 2002, 2004
  - Brute force method
- *ESU/RAND-ESU* – 2006
  - implements *Nauty*
  - subgraph size restricted to 8

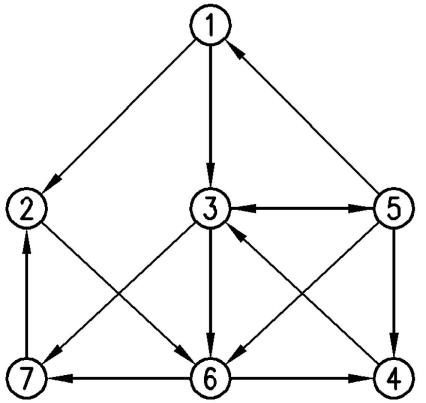


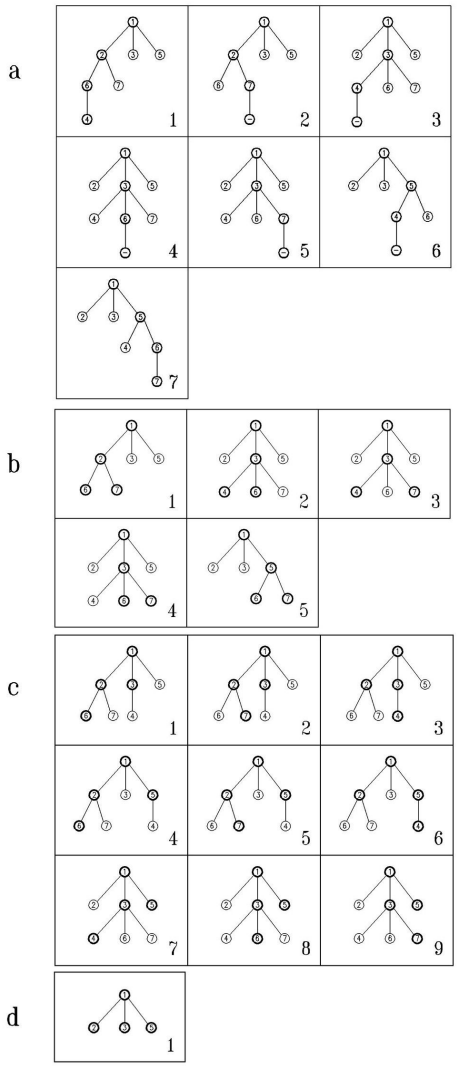
ESU-tree


Brendan D McKay et al. Practical graph isomorphism.
Kashtan et al.: Efficient sampling algorithm for estimating sub-graph concentrations and detecting network motifs.
Wernicke S.: Efficient detection of network motifs.

# Motif discovery – Kavosh - 2009

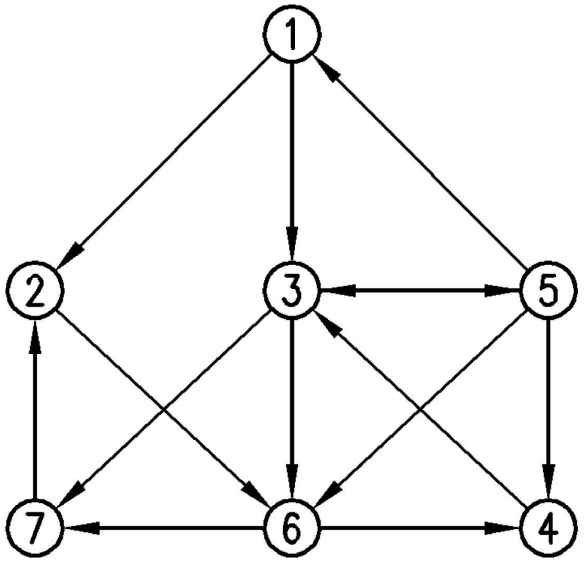
- Iterates on each node.
- Build a tree with max depth *k.*
- Looks at all compositions of *k-1*
- Isomorphic test with *Nauty*



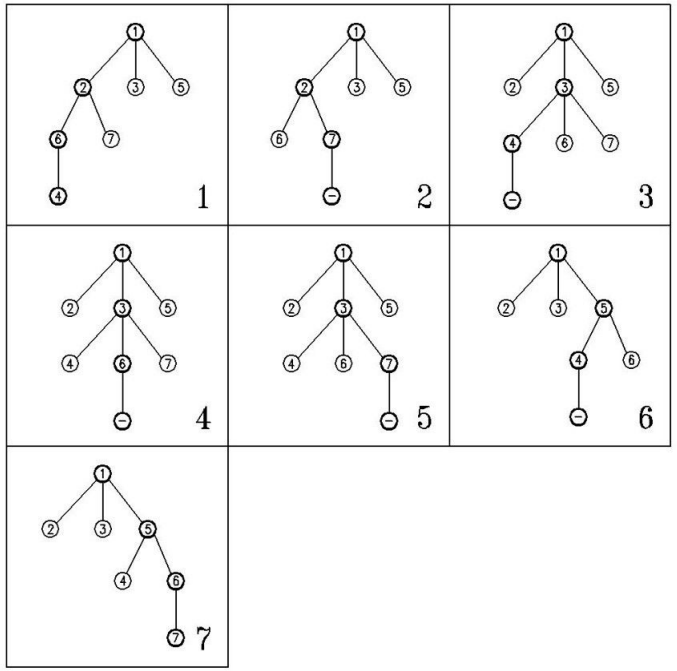




Kashani et al.: Kavosh: a new algorithm for finding network motifs.

# Motif discovery - Kavosh



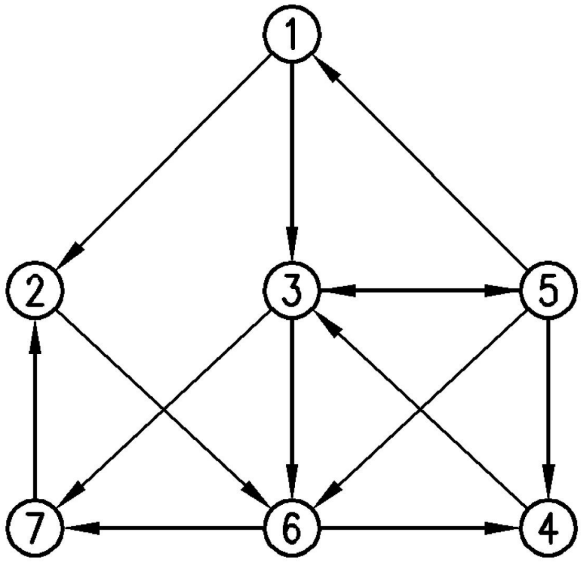

a



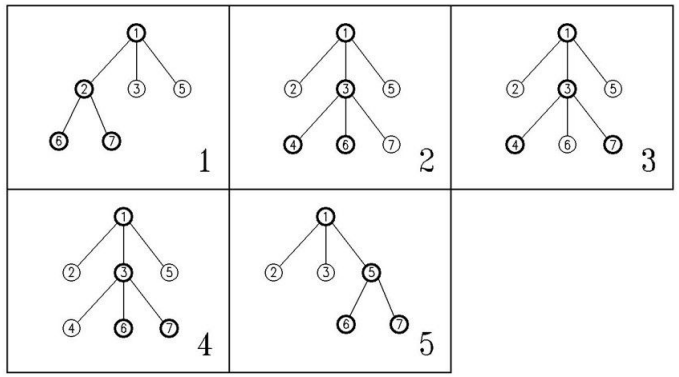

**{ 1 , 1 , 1 , 1 }**

Kashani et al.: Kavosh: a new algorithm for finding network motifs.

# Motif discovery - Kavosh



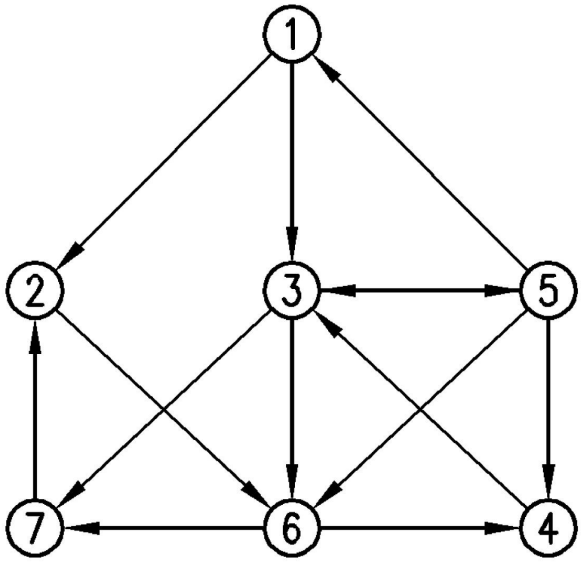

b



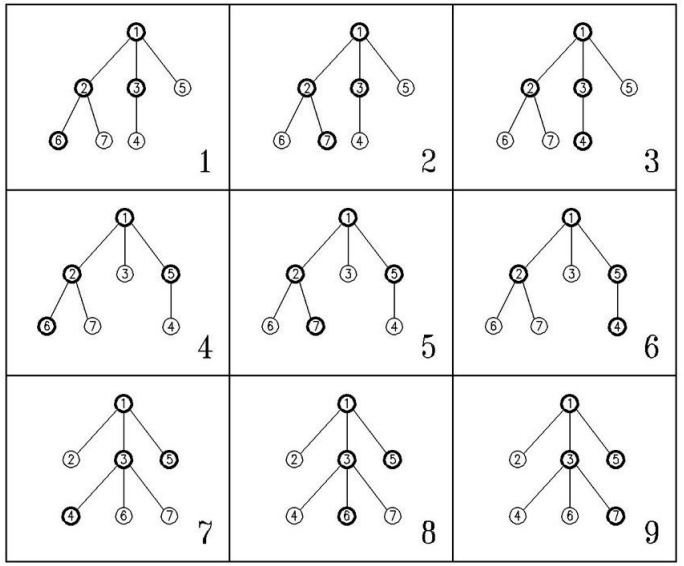

**{1,1,2,0}**

Kashani et al.: Kavosh: a new algorithm for finding network motifs.

# Motif discovery - Kavosh



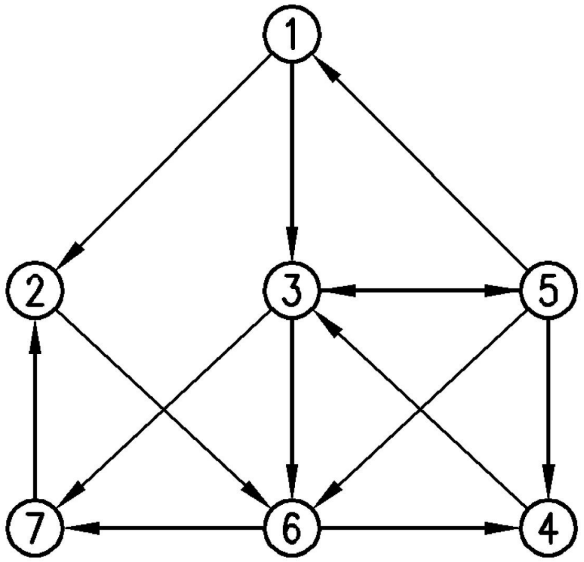

c

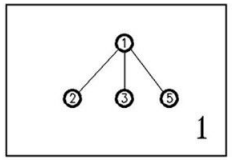



{1,2,1,0}

Kashani et al.: Kavosh: a new algorithm for finding network motifs.

# Motif discovery - Kavosh




d  **{1,3,0,0}**

Kashani et al.: Kavosh: a new algorithm for finding network motifs.

# Motif discovery - Kavosh

- Pros:

  Optimized memory usage

  No subgraph size limit

- Cons:

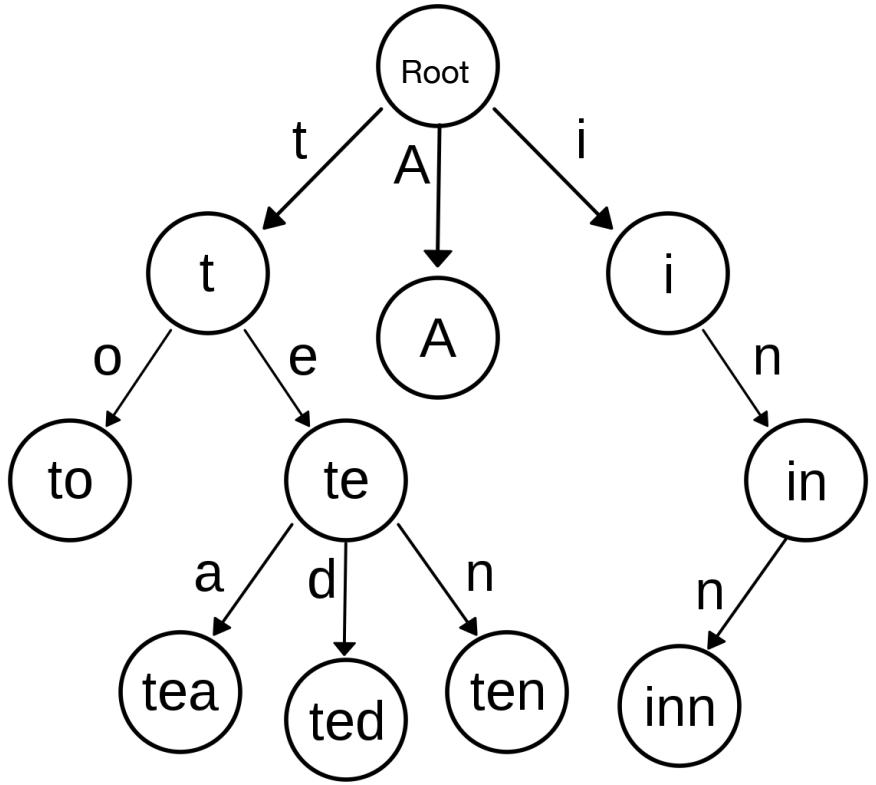
  Not optimized for random network subgraph enumeration



Kashani et al.: Kavosh: a new algorithm for finding network motifs.

# Motif discovery – GtrieScanner - 2010

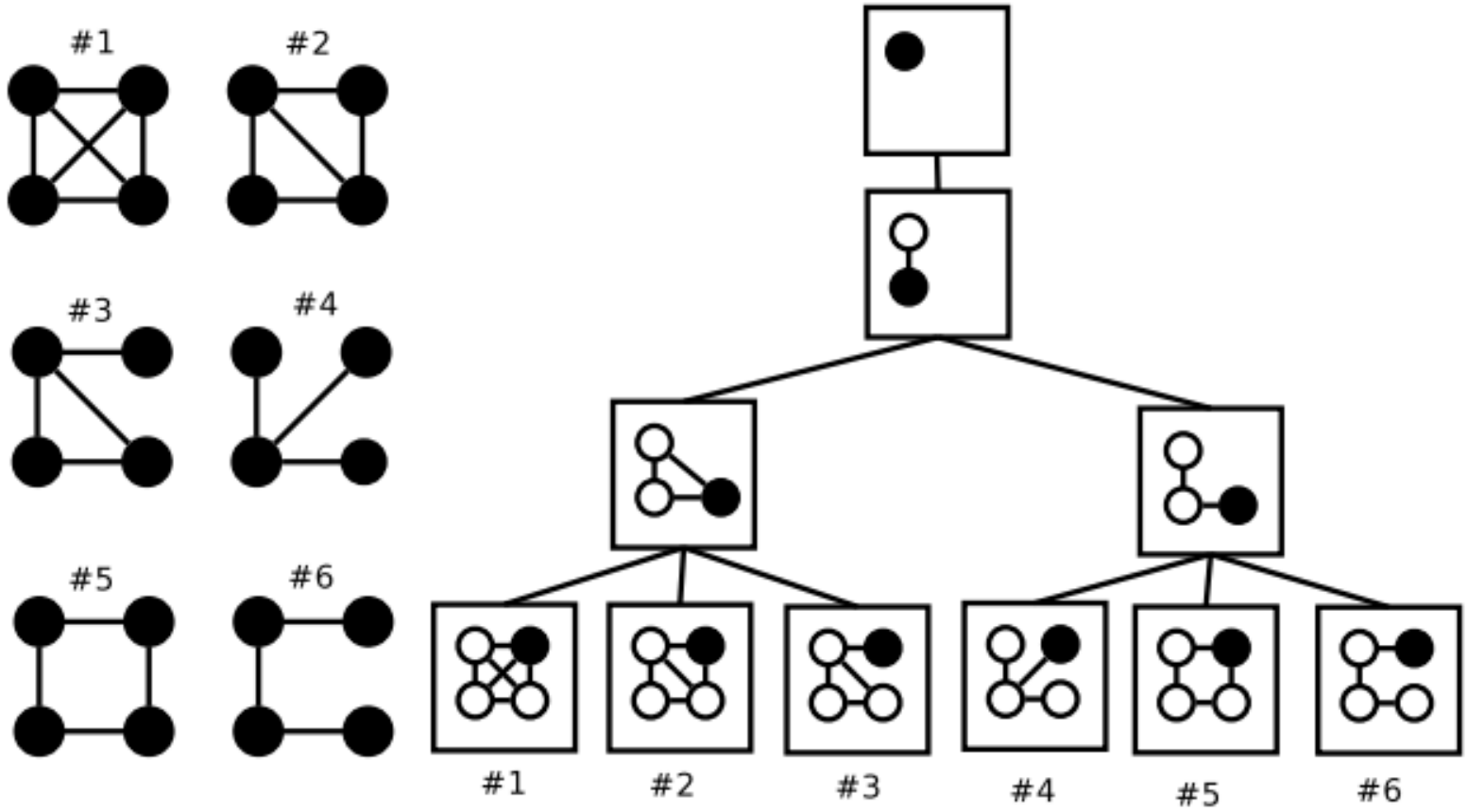
- Gtrie : Prefix tree structure
- First enumeration to build Trie with ESU
- Canonically labelled subgraph adjacency matrices






P. Ribeiro et al.: G-tries: an efficient data structure for discovering network motifs.

# Motif discovery - GtrieScanner




Gtrie for undirected motifs of size 4



P. Ribeiro et al.: G-tries: an efficient data structure for discovering network motifs.

# Motif discovery - GtrieScanner

- Pros:

  Optimized for random network subgraph enumeration

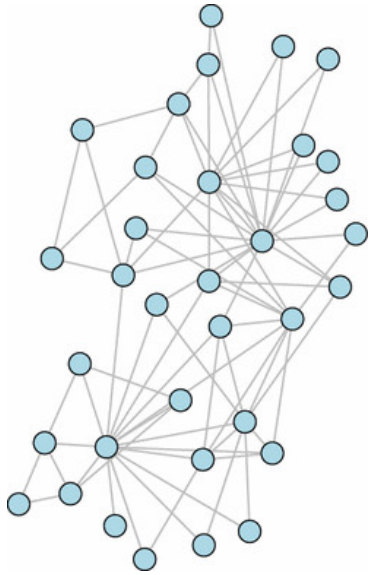
- Cons:

  Need to build Gtrie



P. Ribeiro et al.: G-tries: an efficient data structure for discovering network motifs.

# Motif discovery – Random network generation

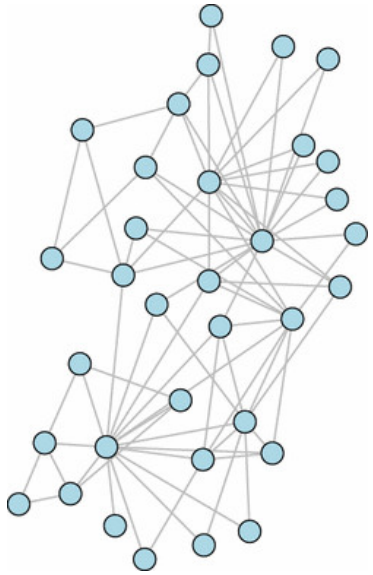


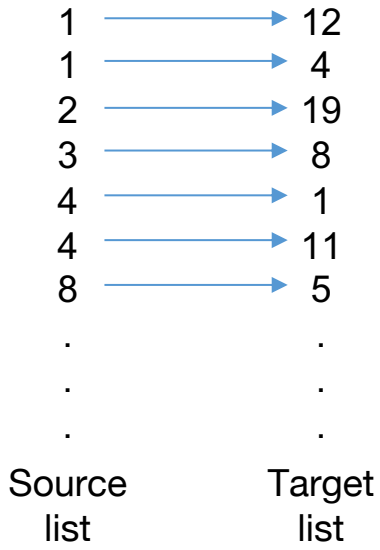
| Source list | Target list |
|---|---|
| 1 | 12 |
| 1 | 4 |
| 2 | 19 |
| 3 | 8 |
| 4 | 1 |
| 4 | 11 |
| 8 | 5 |
| . | . |
| . | . |
| . | . |

Preserve in and out degree

| Source list | Target list |
|---|---|
| 1 | 12 |
| 1 | 4 |
| 2 | 19 |
| 3 | 8 |
| 4 | 1 |
| 4 | 11 |
| 8 | 5 |
| . | . |
| . | . |
| . | . |

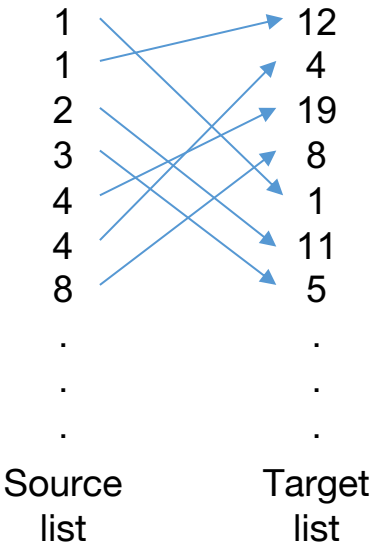
# Motif discovery – Random network generation





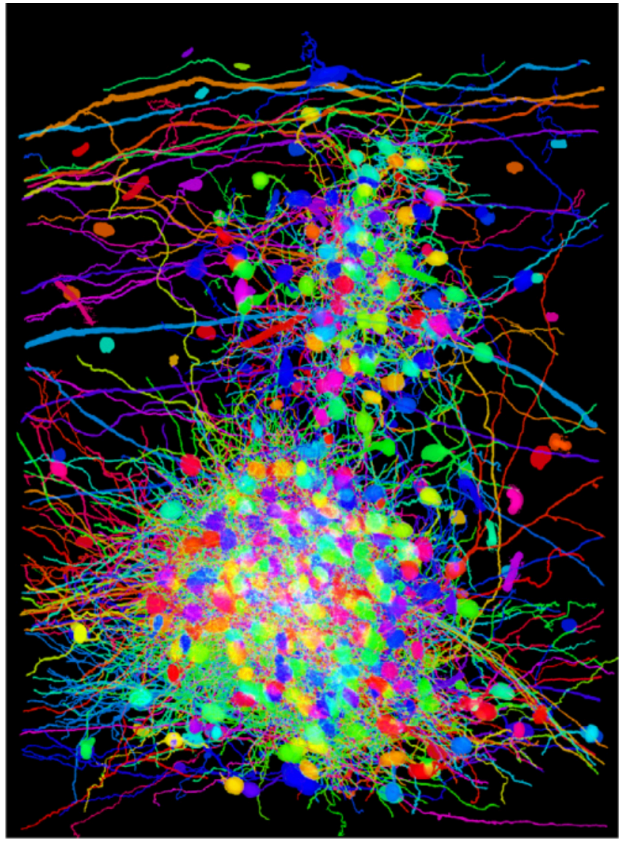

Preserve in and out degree

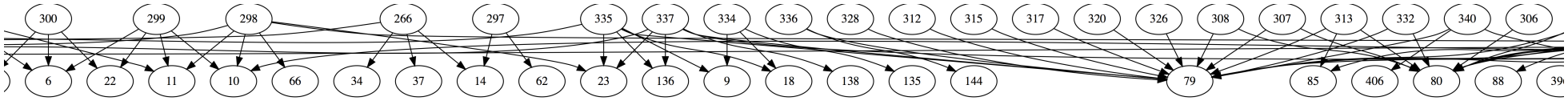
Assign random source and target



# Experiments – LGN dataset

- Mouse dorsal Lateral Geniculate Nucleus
- 400x600x280$\mu m^3$
- 412 nodes
- 820 directed edges






J. Morgan. et al.: The fuzzy logic of network connectivity in mouse visual thalamus.

# Experiments – Fib25 dataset

- Medulla Oblongata of fruit fly (64x66x80$\mu m^3$)
- Initially 798 neuron segments and 31000 synapses
- Removed edges with less than 10 connections
- 279 nodes
- 441 directed edges



S. Takemura et al.: Synaptic circuits and their variations within different columns in the visual system of drosophila.

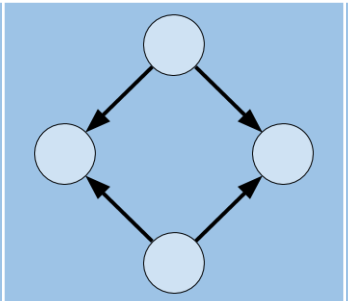
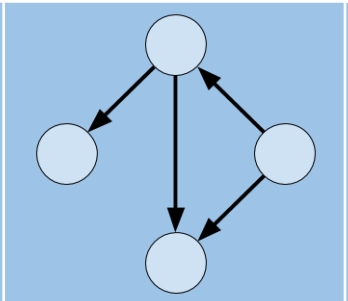
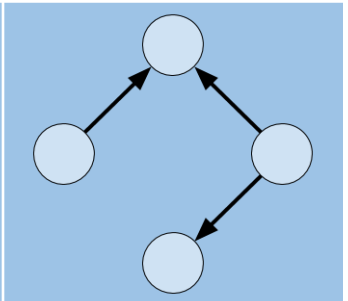
# Survey and application

- Small subgraph enumeration
  - Motif relevance
  - Computing performance

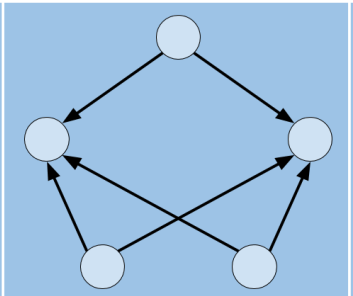
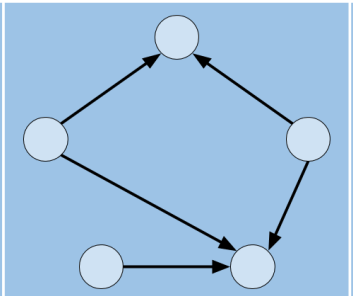
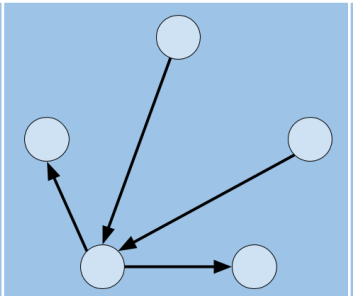
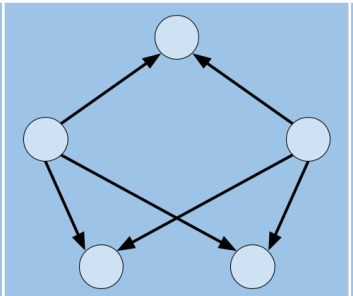
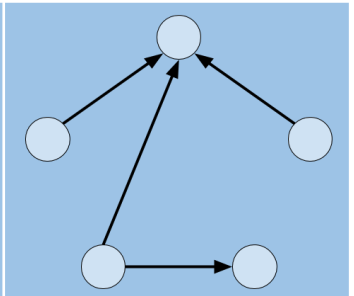
# Results – Small motif relevance

| | LGN | | Fib-25 | |
|---|---|---|---|---|
| Motif Size | Number of Subgraphs | Number of iso. classes | Number of Subgraphs | Number of iso. classes |
| 4 | 89 641 | 18 | 17 239 | 91 |
| 5 | 1 289 357 | 85 | 157 160 | 666 |

# Results – Small motif relevance

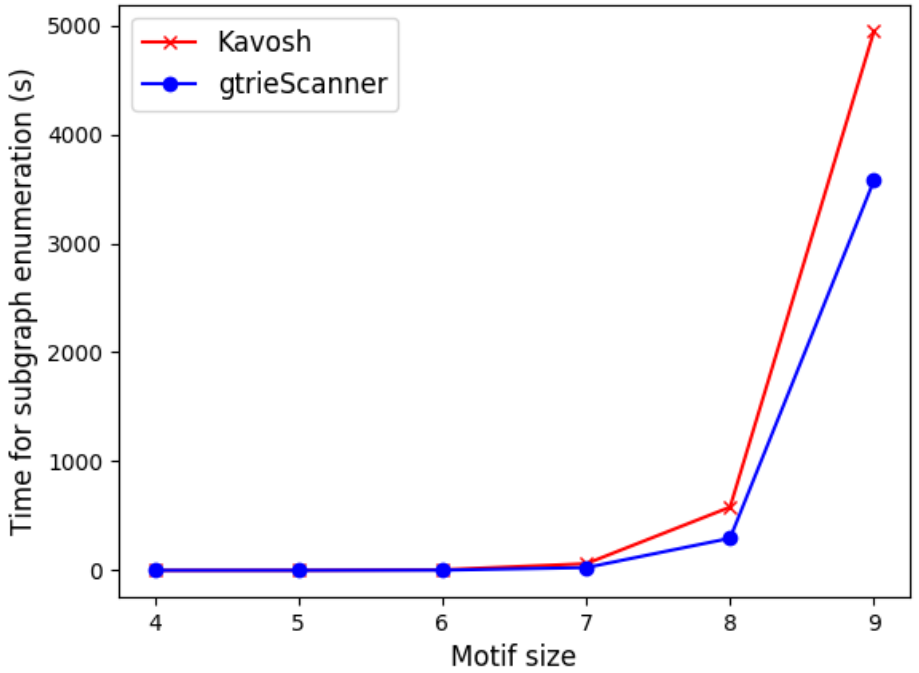
| | Motif ID | 4.1 | 4.2 | 4.3 |
|---|---|---|---|---|
| | Motif |  |  |  |
| LGN | Normalized frequency | 1.8% | 0.15% | 42% |
| | Z-Score | 7.7 | 4.0 | -10.9 |
| Fib-25 | Normalized frequency | 0.68% | 1.6% | 18% |
| | Z-Score | 14.1 | 5.2 | -3.9 |

# Results – Small motif relevance

| | Motif ID | 5.1 | 5.2 | 5.3 | 5.4 | 5.5 |
|---|---|---|---|---|---|---|
| | Motif |  |  |  |  |  |
| LGN | Normalized frequency | 0.15% | 3.0% | 0.04% | 0.08% | 37% |
| | Z-Score | 1.4 | 1.1 | -2.3 | 14.6 | -6.9 |
| Fib-25 | Normalized frequency | 0.06% | 0.79% | 6.3% | 0.01% | 6.3% |
| | Z-Score | 52.1 | 10.7 | -0.7 | 9.4 | -2.7 |

# Results – Small motif computing performance

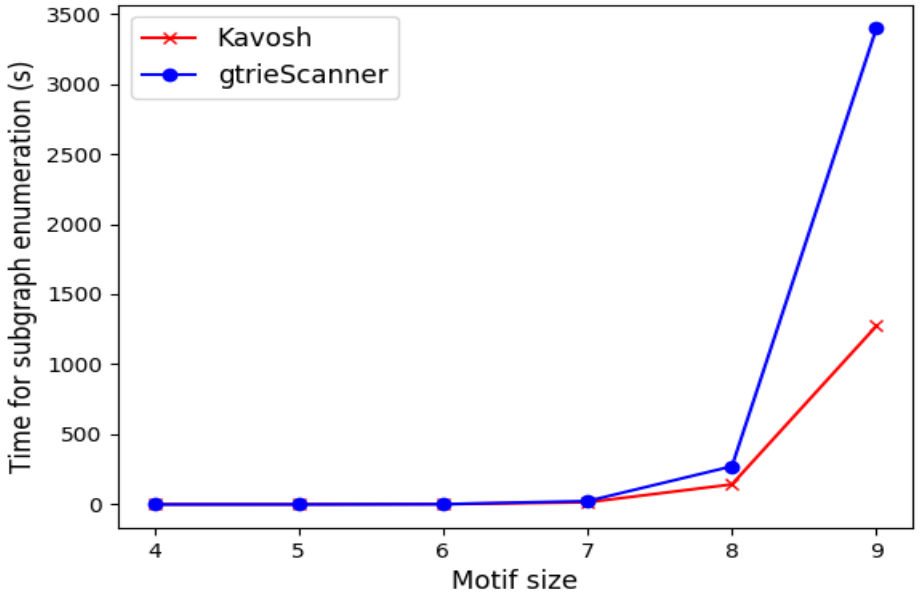
Exponential increase of runtime for both algorithms




Total running time on Fib-25 with 1000 random networks

# Results – Small motif computing performance

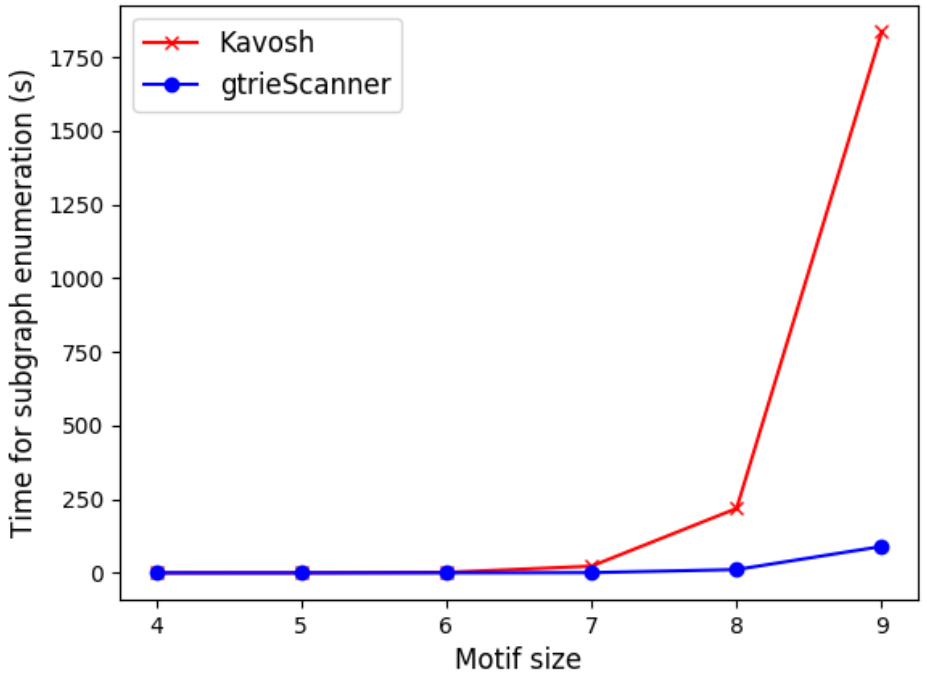
Kavosh is much faster for one graph enumeration




Running time on Fib-25 for the original network

# Results – Small motif computing performance

GtrieScanner is very quick to enurate random graphs

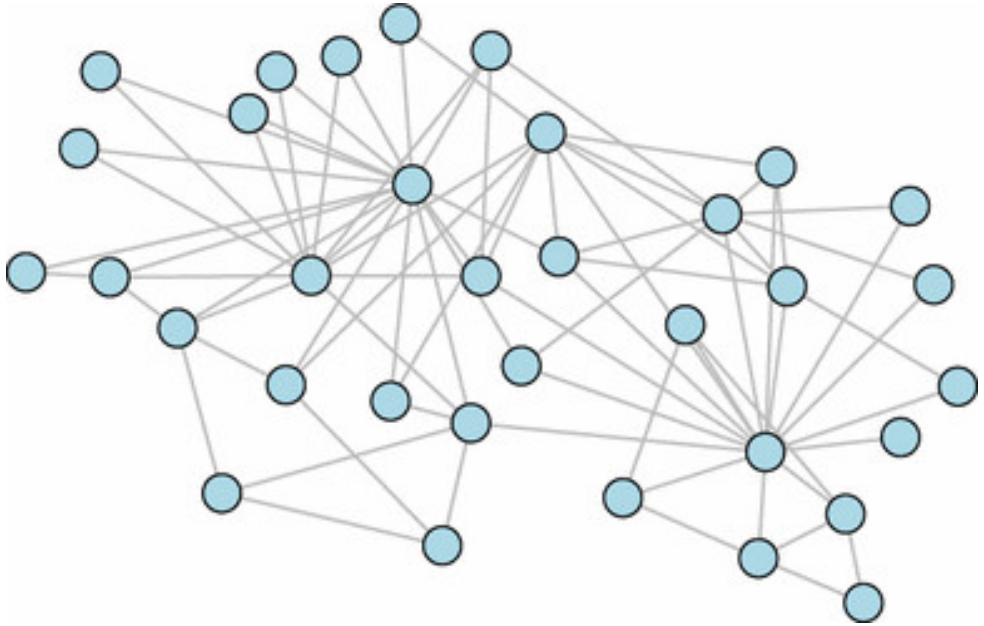



Running time on Fib-25 for 1000 random networks

# Large Subgraph enumeration: Exploration

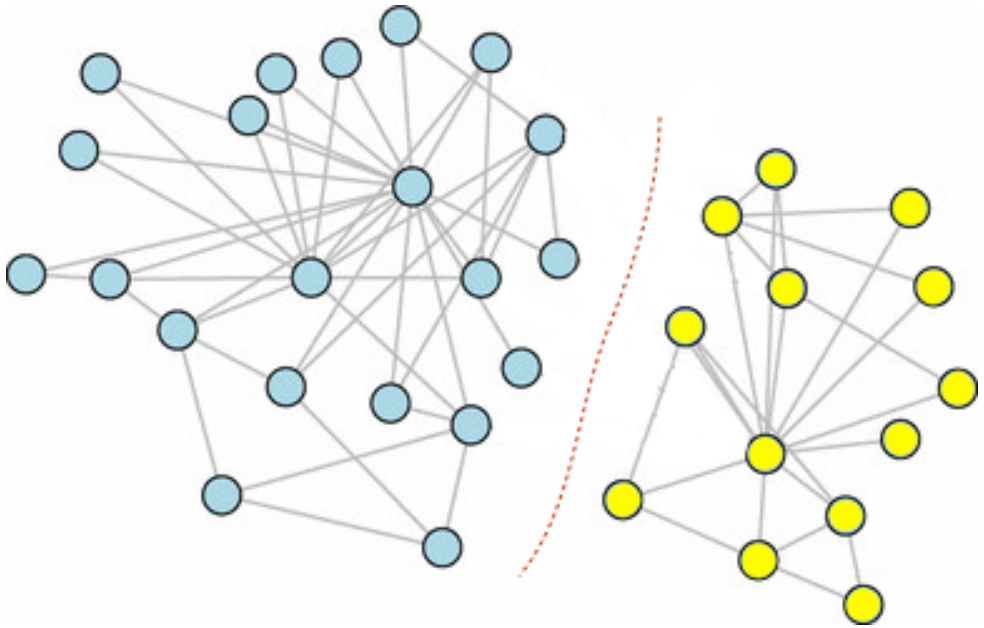
# Methods – Graph theory

Graph cut - $C(V) \in 1, \ldots, n_c$



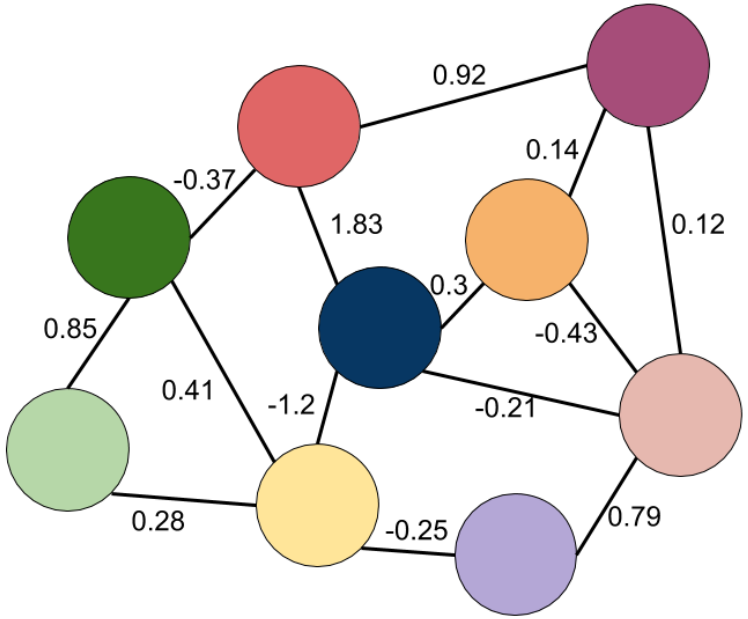
# Methods – Graph theory

Graph cut - $C(V) \in 1, \ldots, n_c$



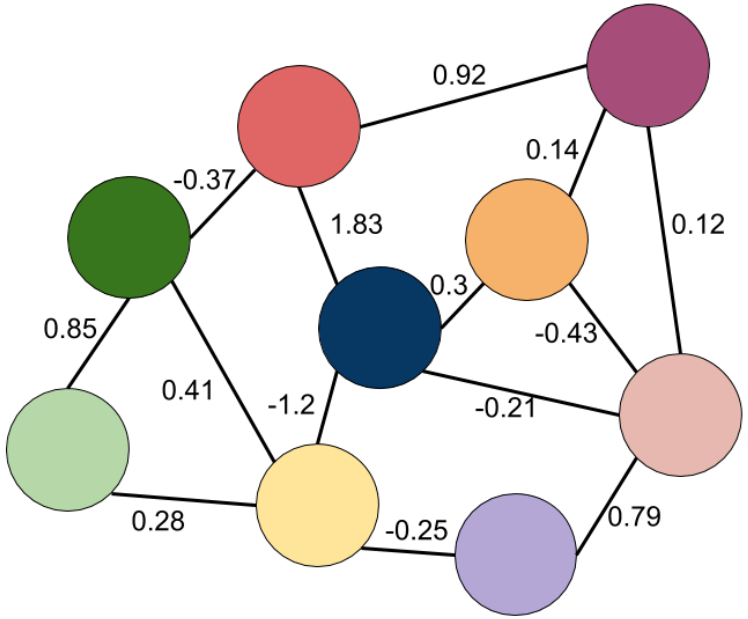
# Motif discovery – Multicut clustering

- Minimum Cost Multicut Problem
- NP hard problem



# Motif discovery – Multicut clustering

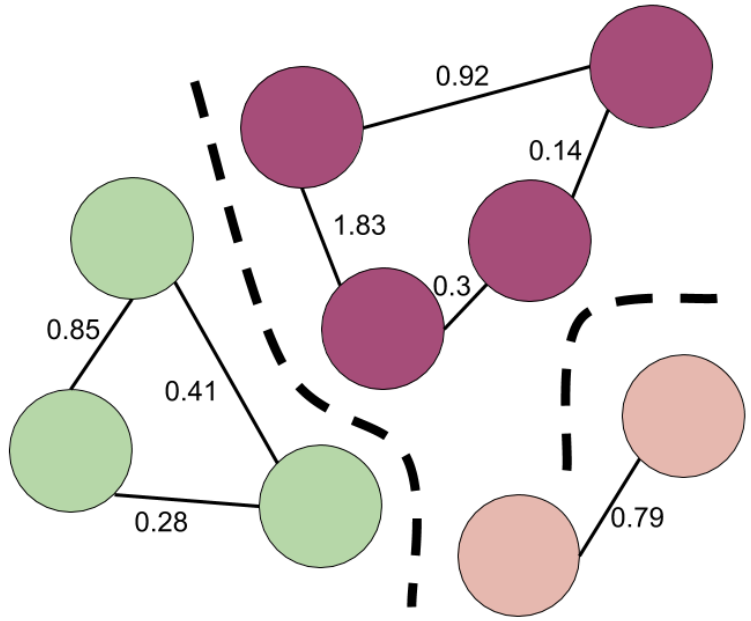
- Minimum Cost Multicut Problem
- NP hard problem
- Modified Kernighan-Lin algorithm
- Iterative algorithm to minimize cost function




M. Keuper et al.: Efficient decomposition of image and mesh graphs by lifted multicuts.

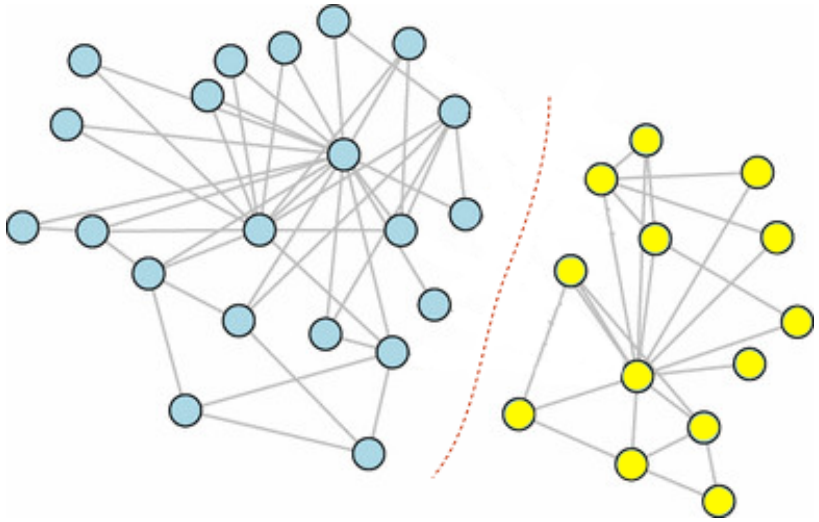
# Motif discovery – Multicut clustering

- Minimum Cost Multicut Problem
- NP hard problem
- Modified Kernighan-Lin algorithm
- Iterative algorithm to minimize cost function
- No need to know number of clusters
- No need for sources



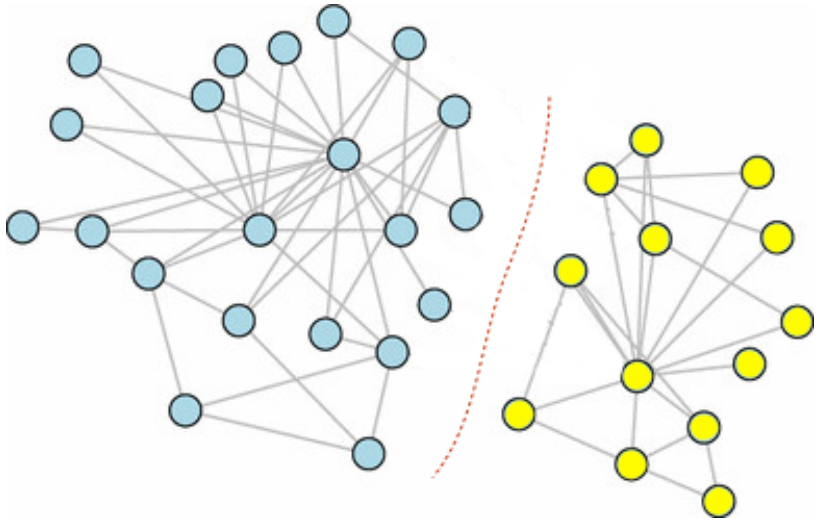

M. Keuper et al.: Efficient decomposition of image and mesh graphs by lifted multicuts.

# OUR PROPOSAL

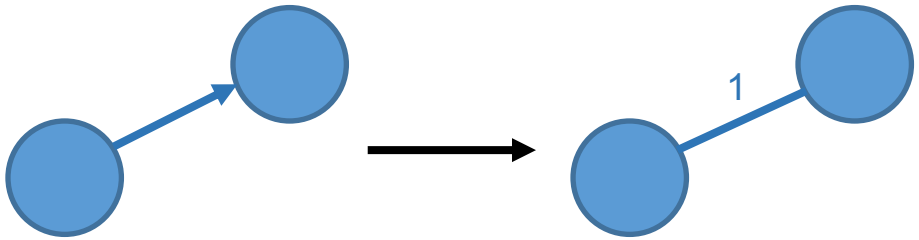


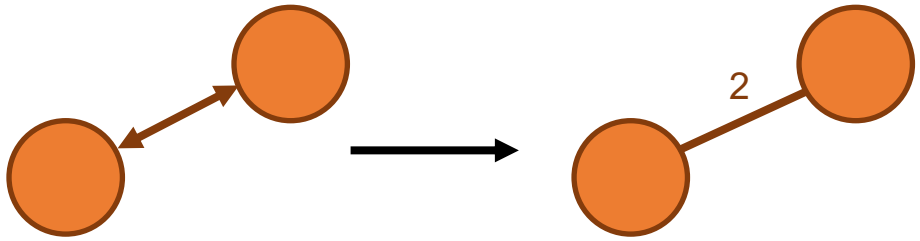
1. Cluster graph

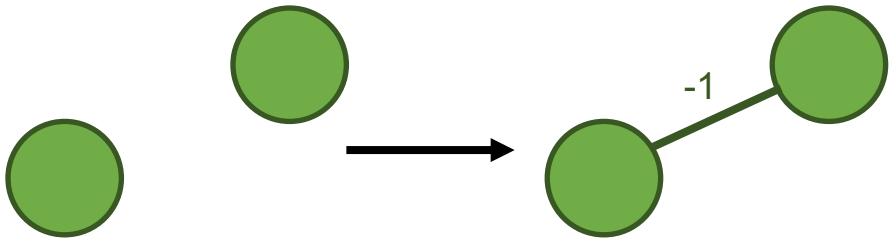
# OUR PROPOSAL



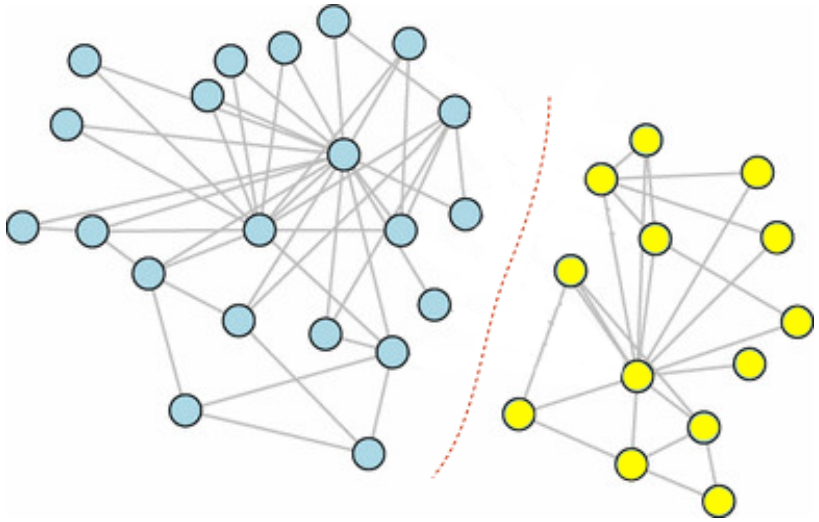
1. Cluster graph



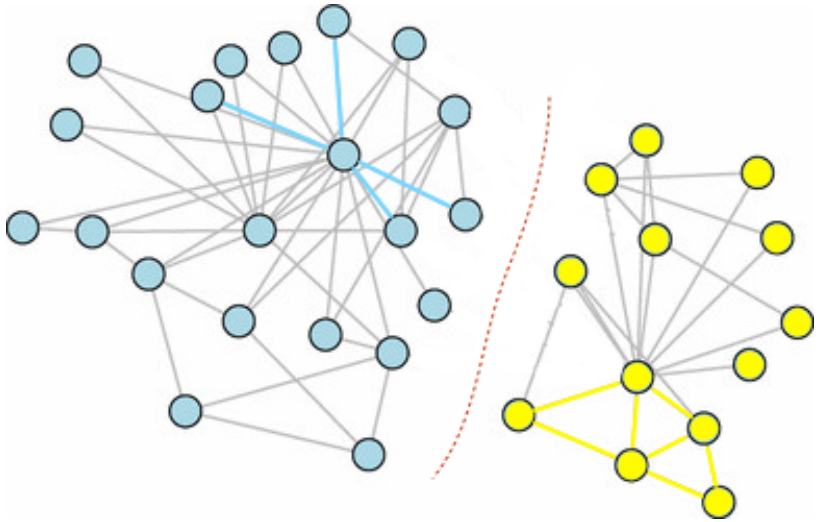



# OUR PROPOSAL



1. Cluster graph



2. Enumerate subgraphs

# Exploration

- Large subgraph enumeration
  - Subgraph frequency
  - Computing performance

# Results – Large motif clustering on *LGN*

| Cluster ID | Number of nodes | Number of edges |
|---|---|---|
| 0 | 200 | 345 |
| 1 | 38 | 45 |
| 2 | 20 | 19 |
| 3 | 5 | 4 |
| 5 | 10 | 7 |
| 8 | 10 | 9 |
| 10 | 36 | 43 |
| 15 | 4 | 3 |
| 16 | 78 | 70 |

# Results – Large motif clustering on *LGN*

| Cluster ID | Number of nodes | Number of edges |
|---|---|---|
| 0 | 200 | 345 |
| 1 | 38 | 45 |
| 2 | 20 | 19 |
| 3 | 5 | 4 |
| 5 | 10 | 7 |
| 8 | 10 | 9 |
| 10 | 36 | 43 |
| 15 | 4 | 3 |
| 16 | 78 | 70 |

# Results – Large motif relevance on *LGN*

- 3 622 227 359 subgraphs occurrences in original graph
- 1 633 037 684 subgraphs occurrences in clustered graph
- 55% search space reduction

# Results – Large motif relevance on *LGN*

| Motif ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Z-Score ↑ | 5100 | 307 | 115 | 83.3 | 72.8 | 56.7 | 39.3 | 22.9 |
| Original frequency ↑ | 510 | 2 385 | 1 945 | 32 235 | 8 415 | 10 055 | 49 182 | 186 620 |
| Clusters frequency | 510 | 2 385 | 1 945 | 32 235 | 8 415 | 10 055 | 465 | 186 620 |

# Results – Large motif relevance on *LGN*

| Motif ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Z-Score ↑ | 5100 | 307 | 115 | 83.3 | 72.8 | 56.7 | 39.3 | 22.9 |
| Original frequency ↑ | 510 | 2 385 | 1 945 | 32 235 | 8 415 | 10 055 | 49 182 | 186 620 |
| Clusters frequency | 510 | 2 385 | 1 945 | 32 235 | 8 415 | 10 055 | 465 | 186 620 |

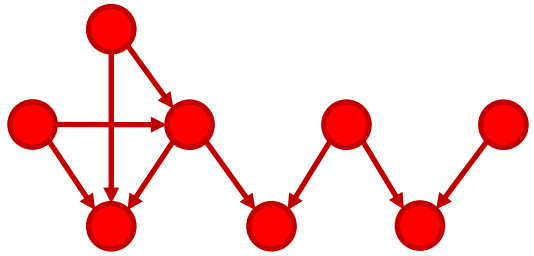
# Results – Large motif relevance on *LGN*

| Motif ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Z-Score ↑ | 5100 | 307 | 115 | 83.3 | 72.8 | 56.7 | 39.3 | 22.9 |
| Original frequency ↑ | 510 | 2 385 | 1 945 | 32 235 | 8 415 | 10 055 | 49 182 | 186 620 |
| Clusters frequency | 510 | 2 385 | 1 945 | 32 235 | 8 415 | 10 055 | 465 | 186 620 |

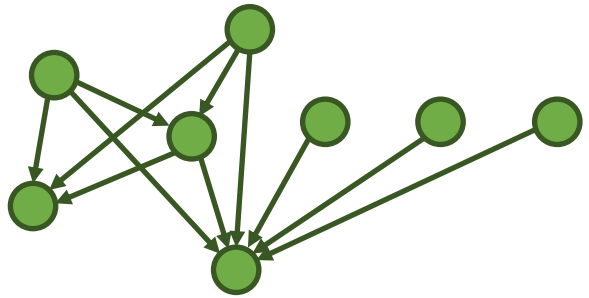
# Results – Large motif relevance on *LGN*

| 7 | 8 |
|---|---|
| 39.3 | 22.9 |
| 49 182 | 186 620 |
| 465 | 186 620 |

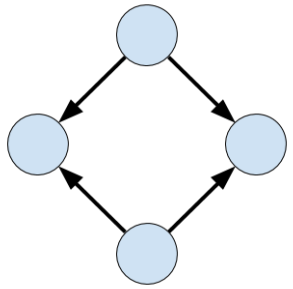
```
01001000
00000000
11000000
11000000
00000000
00001010
00000000
00000010
```



```
00000000
10001000
11001000
11001000
00000000
10000000
10000000
10000000
```



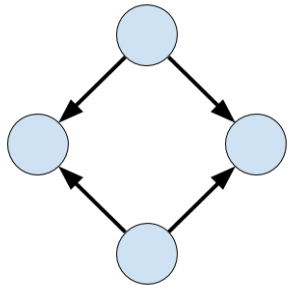
# Results – Large motif relevance on *LGN*



0011
0011
0000
0000

Motif 4.1

# Results – Large motif relevance on *LGN*

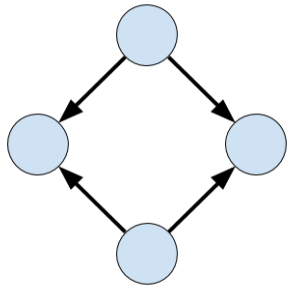


0011
0011
0000
0000

Motif 4.1

| Motif 1 | Motif 2 | Motif 3 | Motif 4 | Motif 5 | Motif 6 | Motif 7 | Motif 8 |
|---|---|---|---|---|---|---|---|
| 00110000 | 00110000 | 00110000 | 00110000 | 00000000 | 00110001 | 01001000 | 00000000 |
| 10110010 | 10110000 | 10110000 | 10110000 | 10001000 | 10110000 | 00000000 | 10001000 |
| 00000000 | 00000000 | 00000001 | 00000000 | 11001100 | 00000000 | 11000000 | 11001000 |
| 00000000 | 00000000 | 00000000 | 00000000 | 11001000 | 00000000 | 11000000 | 11001000 |
| 10110000 | 10110000 | 10110000 | 10110000 | 00000000 | 10110000 | 00000000 | 00000000 |
| 10010010 | 10110000 | 10110000 | 10110000 | 10000000 | 10110000 | 00001010 | 10000000 |
| 00100000 | 10110000 | 00100000 | 00100000 | 10000000 | 00100000 | 00000000 | 10000000 |
| 00100000 | 00100000 | 00000000 | 00100000 | 10000000 | 00000000 | 00000010 | 10000000 |

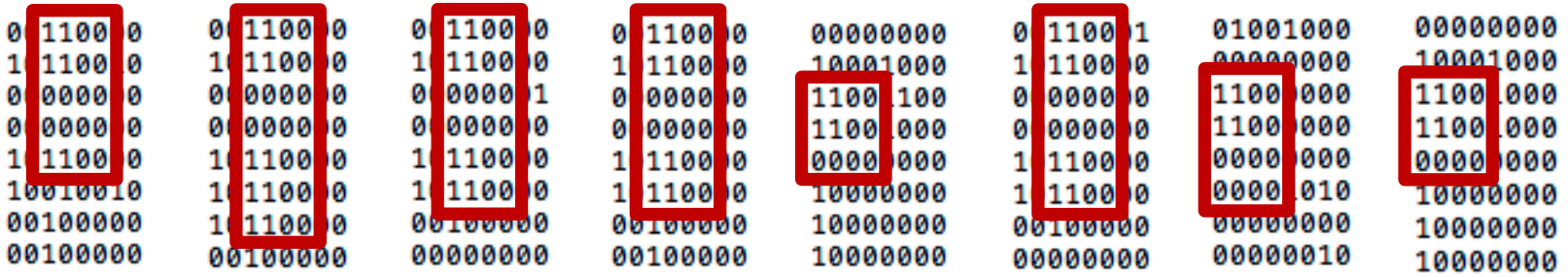
# Results – Large motif relevance on *LGN*



Motif 4.1

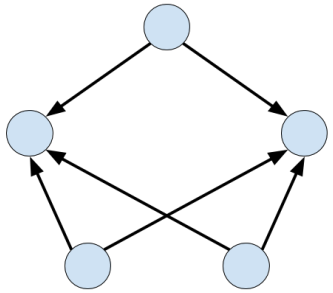
0011
0011
0000
0000

Identified in all size 8 motifs of LGN dataset



Motif 1 Motif 2 Motif 3 Motif 4 Motif 5 Motif 6 Motif 7 Motif 8

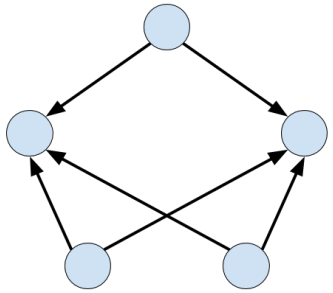
# Results – Large motif relevance on *LGN*



Motif 5.1

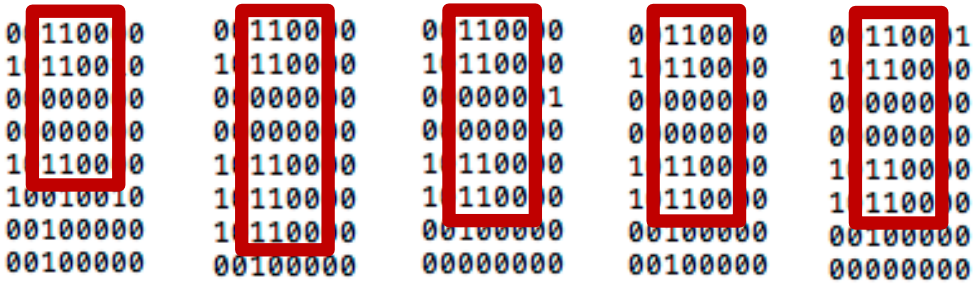
00011
00011
00011
00000
00000

# Results – Large motif relevance on *LGN*



Motif 5.1

00011
00011
00011
00000
00000



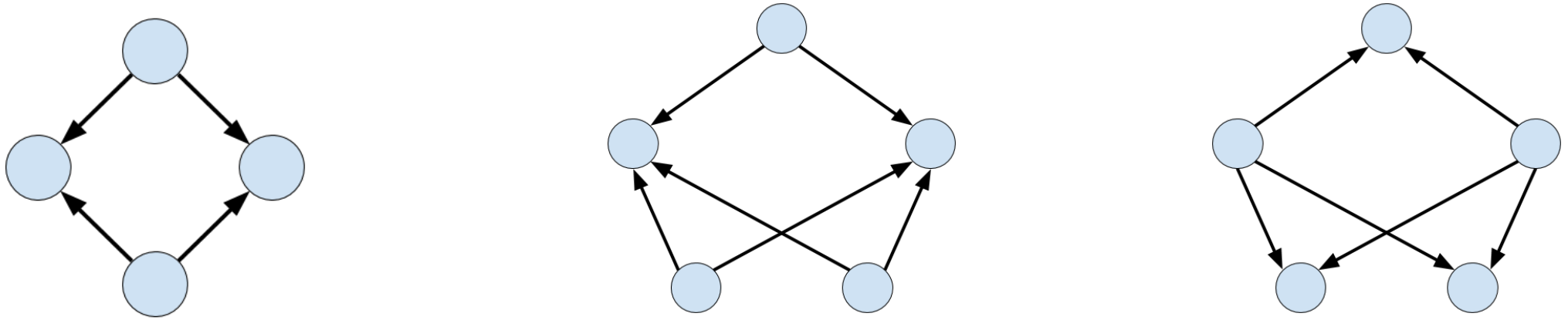
Motif 1 Motif 2 Motif 3 Motif 4 Motif 6

# Results – Large motif performance on *LGN*

| | GtrieScanner | OUR PROPOSAL |
|---|---|---|
| **Clustering time (s)** | 0 | 4 |
| **Subgraph enumeration time (s)** | 15 860 | 4 821 |
| **Total time (s)** | **15 860 (~4h30m)** | **4 825 (~1h20m)** |

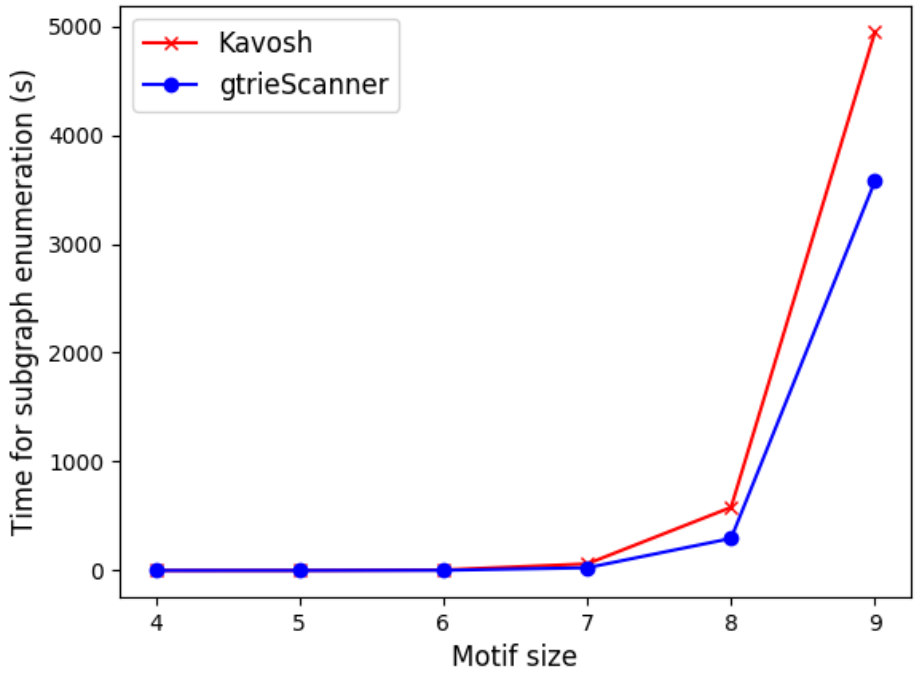
## Discussion

- From the survey, same motifs have been identified.
- Two real connectomes (mouse and fruit fly).



## Discussion

- Kavosh is very efficient to enumerate a single graph
- GtrieScanner is more time efficient when computing subgraph statistics



## Discussion

- We have proposed a solution to speed up subgraph enumeration

3x faster

- Drawbacks:
  - Clusters are mostly uneven.

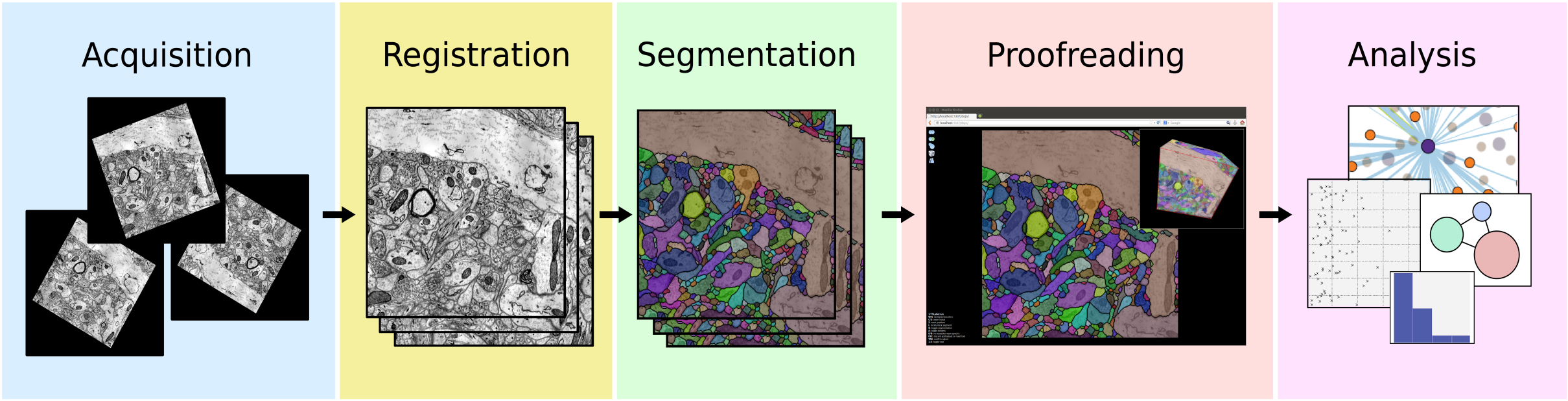
## Future Works

- Improve the clustering on our algorithm.
- Apply small motif discovery to more brain wiring diagram.

**Thank you to the entire VCG lab and to Prof. Kathryn Hess!**

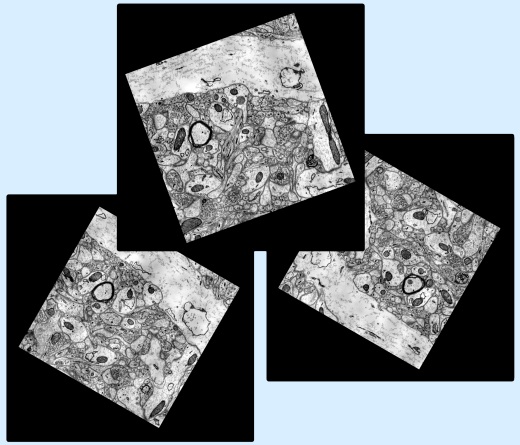
# Thank you

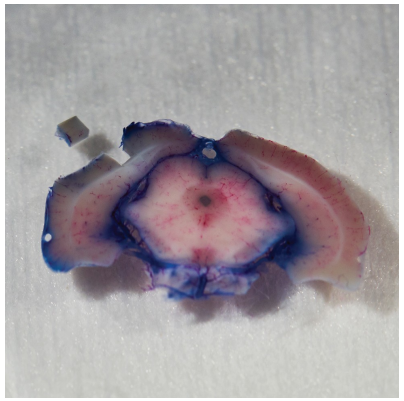
# Connectomics




D. Haehn et al. Scalable interactive visualization for connectomics. *Informatics.*

## Acquisition

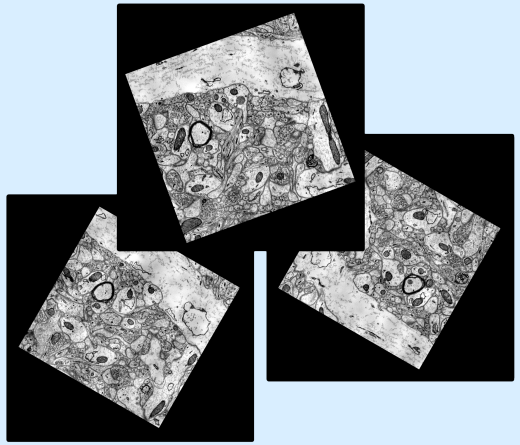


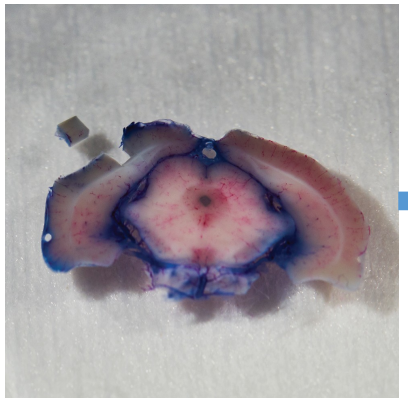


Cubic millimeter of brain tissue.

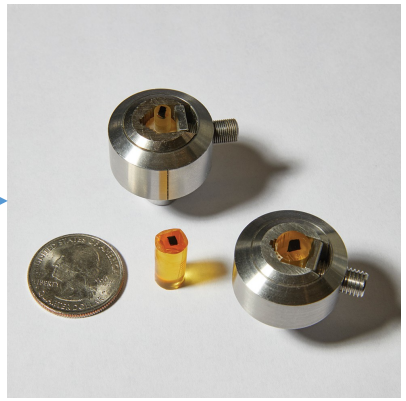
Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

## Acquisition





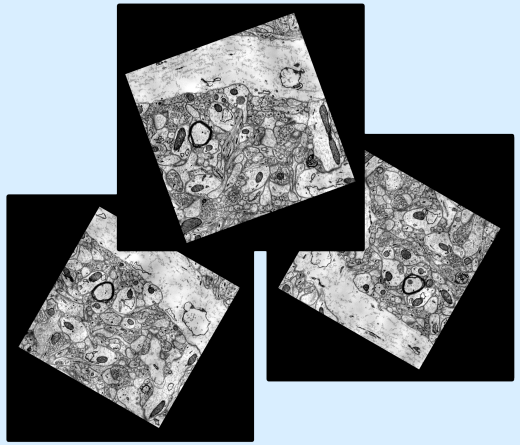
Cubic millimeter of brain tissue.

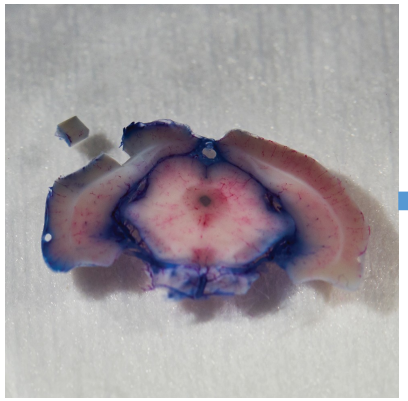


Thinly sliced in an acrylic encasing.

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

## Acquisition





Cubic millimeter of brain tissue.

Thinly sliced in an acrylic encasing.

EM imaging

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

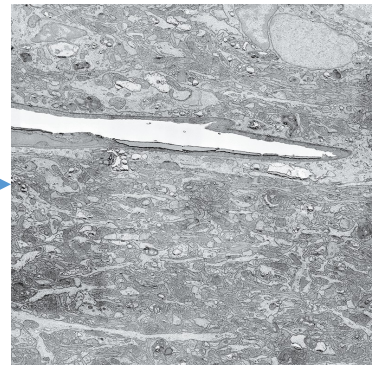
## Registration

EM images…

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

## Registration

EM images…



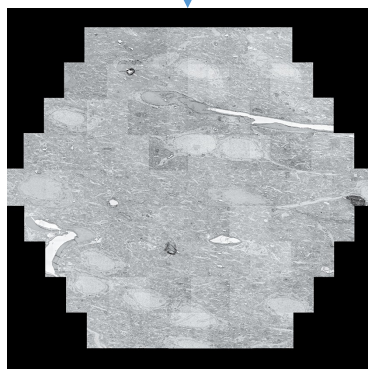
…are aligned together…

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

## Registration

EM images…

…are aligned together…



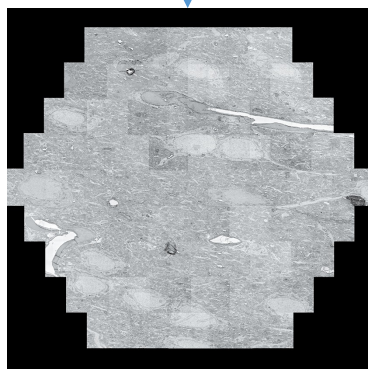
… into fields of views…

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

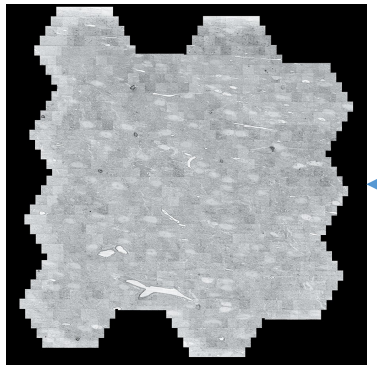
# Registration

EM images…

…are aligned together…



… into fields of views…



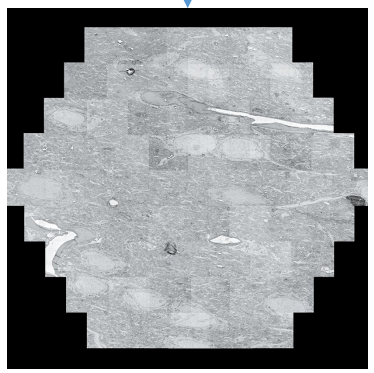
… aligned together to reconstruct a slice…

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

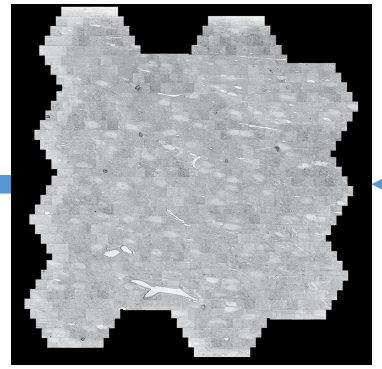
# Registration

EM images…

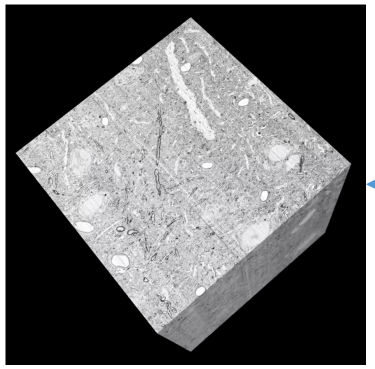
…are aligned together…



… into fields of views…



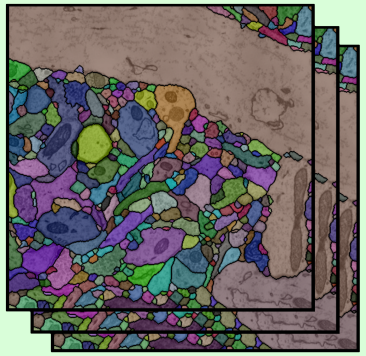
… aligned together to reconstruct a slice…

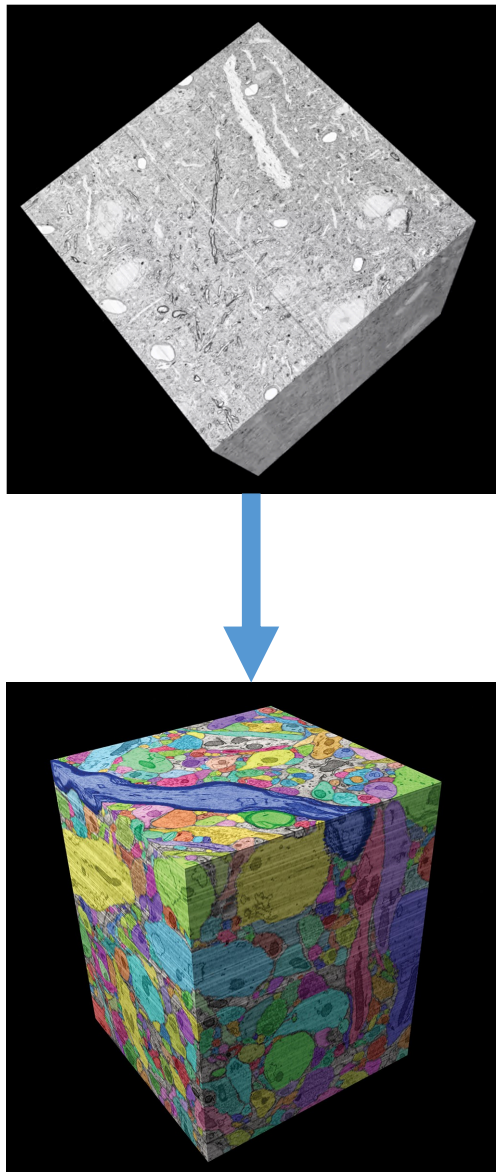


…finally registered into a 3D volume.

Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

## Segmentation





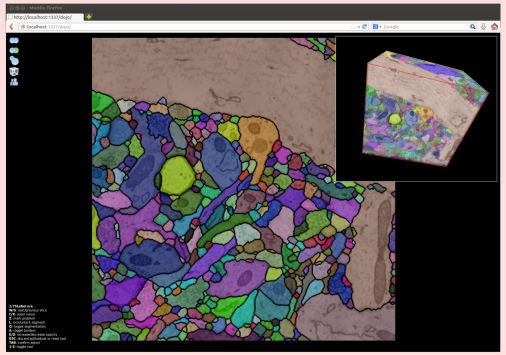
Apply automatic 3D segmentation to registered slices.

Segment neurons and synapses.

Computationally very costly.

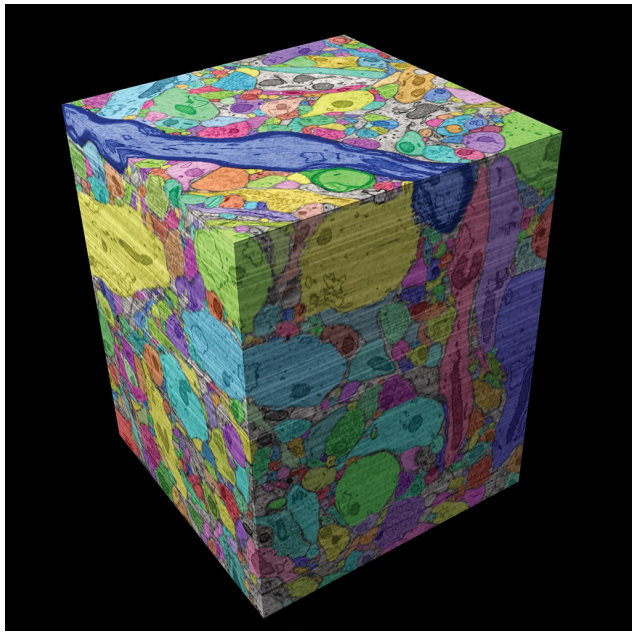
Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

# Proofreading



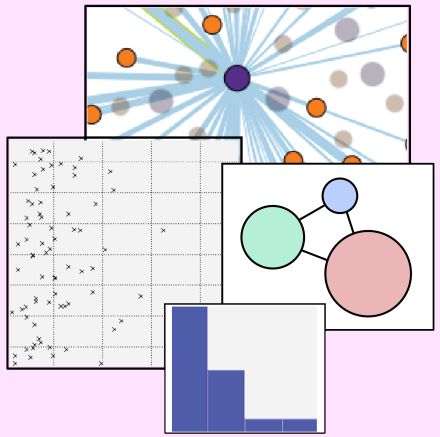
Proofreading is a necessary step in connectomics

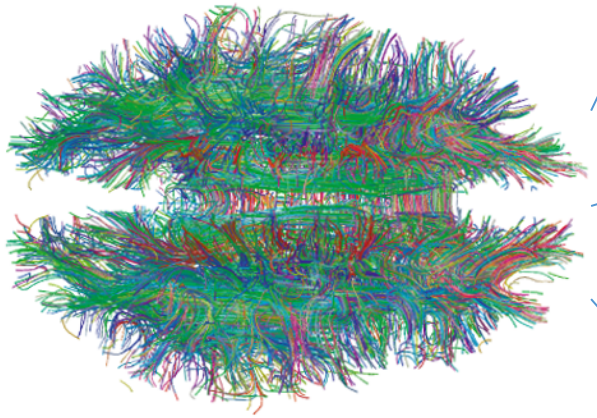
Focuses on streamlining human involvement

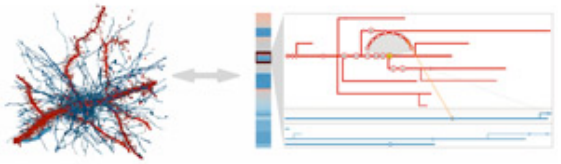


Inside the Moonshot Effort to Finally Figure Out the Brain. M. Mitchell Waldrop (Oct 12, 2017) MIT Technology Review

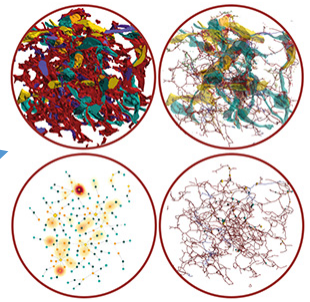
## Analysis



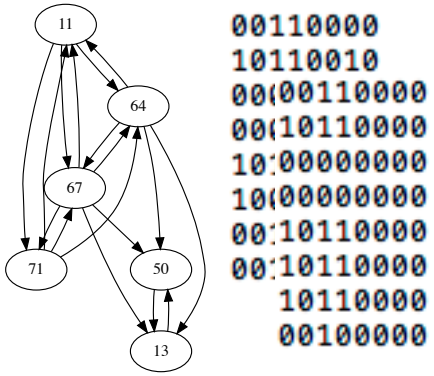




Connectome visualization



Glial cells visualization




Motif Discovery

H. Mohammed et al.: Abstractocyte: A Visual Tool for Exploring Nanoscale Astroglial Cells
A. Al-Awami et al.: NeuroLines: A Subway Map Metaphor for Visualizing Nanoscale Neuronal Connectivity