

# A Computationally Driven Comparative Survey of Network Alignment, Graph Matching, and Network Comparison in Pattern Recognition and Systems Biology

Marissa Graham  
Advisor: Dr. Emily Evans

Brigham Young University

# Motivation

- Networks are a powerful way to model relationships between objects

# Motivation

- Networks are a powerful way to model relationships between objects
- How can we exploit this representation to gain insights about the objects being modeled?

# Motivation

- Networks are a powerful way to model relationships between objects
- How can we exploit this representation to gain insights about the objects being modeled?
  - Visualization
  - Connectivity, paths, centralities, degree distributions, etc.
  - Comparison between networks

# Motivation

- Two main fields of application for network comparison

# Motivation

- Two main fields of application for network comparison
  - Pattern recognition (computer science)
  - Systems biology

# Motivation

- Two main fields of application for network comparison
  - Pattern recognition (computer science)
  - Systems biology
- Existing surveys are not truly interdisciplinary

# Motivation

- Two main fields of application for network comparison
  - Pattern recognition (computer science)
  - Systems biology
- Existing surveys are not truly interdisciplinary
  - Lack of common terminology and background inhibits understanding
  - “Reference dumping”
  - Difficult to determine most relevant works without prior expertise in a field

# Motivation

- Two main fields of application for network comparison
  - Pattern recognition (computer science)
  - Systems biology
- Existing surveys are not truly interdisciplinary
  - Lack of common terminology and background inhibits understanding
  - “Reference dumping”
  - Difficult to determine most relevant works without prior expertise in a field
- How can we create a more cohesive presentation?

# Approach

- Curate a list of relevant papers
  - First five pages of Google Scholar for ten different search terms

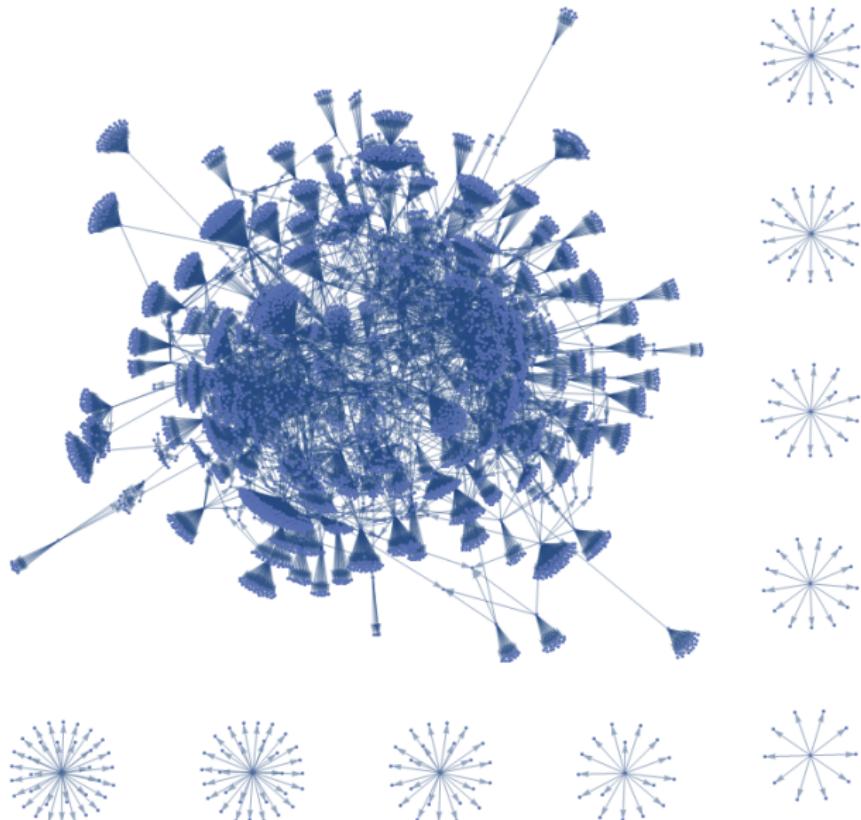
# Approach

- Curate a list of relevant papers
  - First five pages of Google Scholar for ten different search terms
- Construct a citation network from their reference lists

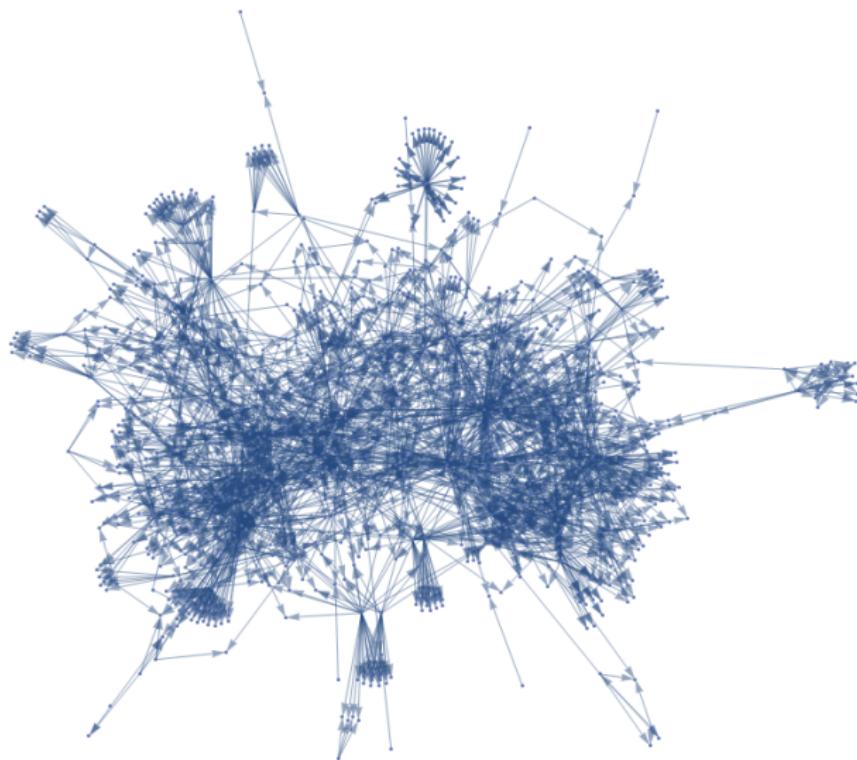
# Approach

- Curate a list of relevant papers
  - First five pages of Google Scholar for ten different search terms
- Construct a citation network from their reference lists
  - Hand-format .txt file reference lists
  - Look up metadata from freeform citation using the CrossRef REST API
  - Manually check unverified results and parse incorrect results

# The full citation network



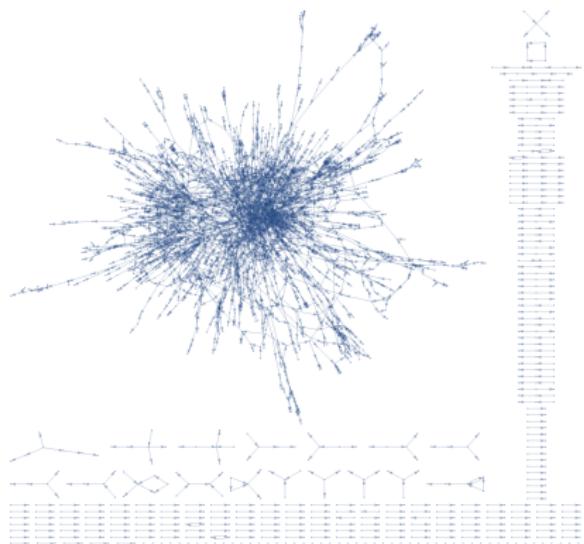
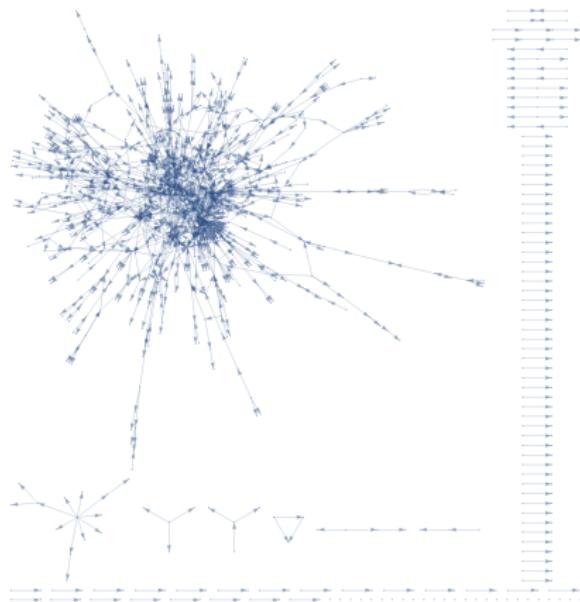
# The pruned citation network



# Statistics

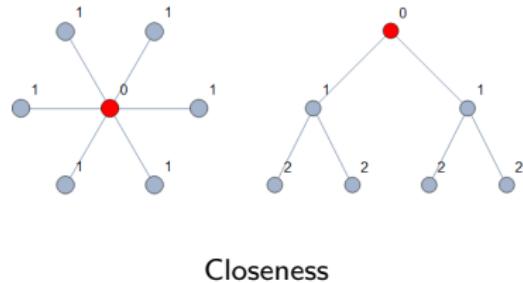
	$G$	$G_p$	sciMet	zewail	$R$	$R_d$
Vertices	5793	1062	1092	3145	5793	5793
Edges	7491	2775	1308	3743	7491	7491
Mean degree	1.29	2.61	1.20	1.19	1.29	1.29
Fraction with children	0.038	0.193	0.523	0.599	0.733	0.038
Diameter	10	9	14	22	21	9
Connected components	16	1	114	281	504	3
% in giant component	96	100	78.4	79.7	90	99.9

# SciMet and Zewail (for reference)



# Centrality

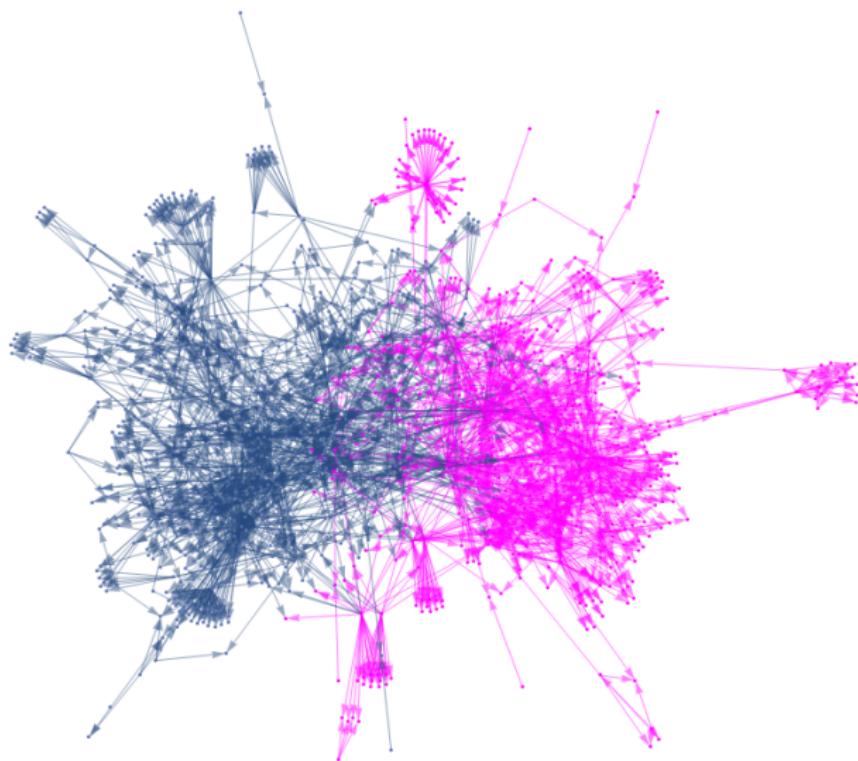
- Indegree
- Outdegree
- Betweenness
- Closeness
- HITS authority
- HITS hub



# Centrality

	Indegree	Outdegree	Betweenness	Closeness	HITS Auth.	HITS Hub
◊ Thirty years of graph matching in pattern recognition [21]	20*	109*	1	2		1
† Fifty years of graph matching, network alignment and network comparison [32]	6	71*	2	1		3
† Networks for systems biology: Conceptual connection of data and function [31]	2	102*	3	3		2
◊ An algorithm for subgraph isomorphism [102]	20*	4	7	4	1	
† Modeling cellular machinery through biological network comparison [95]	9	41*	8			
◊ Computers and intractability: A guide to the theory of NP-completeness [44]	16*	0	4	5		
◊ The graph matching problem [66]	2	55*	5	6		7
† A new graph-based method for pairwise global network alignment [55]	9	13		8		
† On graph kernels: Hardness results and efficient alternatives [41]	11	10	6			
◊ Error correcting graph matching: On the influence of the underlying cost function [10]	10	16		7	7	8
◊ A graduated assignment algorithm for graph matching [39]	18*	0			5	
◊ The Hungarian method for the assignment problem [61]	17*	0				

# Partitioning



# Partitioning

- Do the “clumps” seen match our observations about the two main fields of application?

# Partitioning

- Do the “clumps” seen match our observations about the two main fields of application?
- Subject information not available in metadata

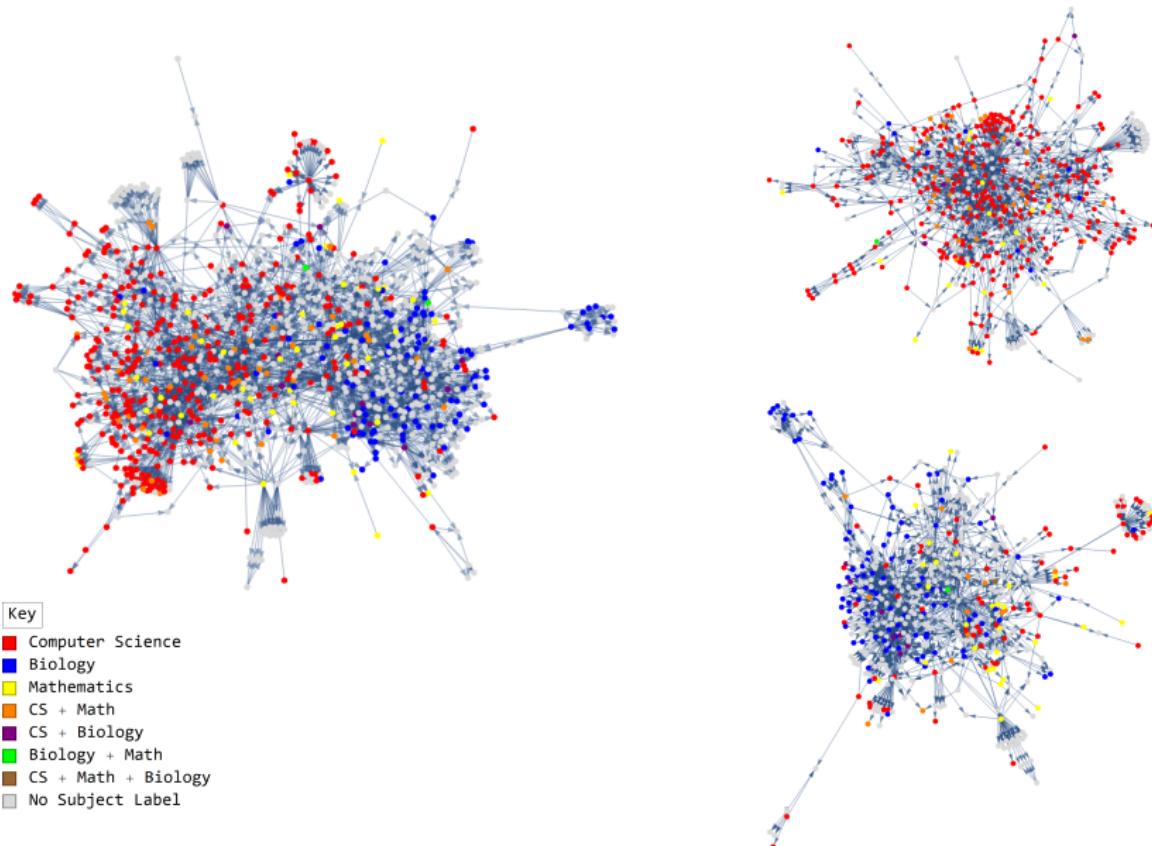
# Partitioning

- Do the “clumps” seen match our observations about the two main fields of application?
- Subject information not available in metadata
- Tag papers using journal titles (which are available)
  - Ex) “biology”, “cell”, “protein”, “DNA”
  - Ex) “computer vision”, “ACM”, “IEEE”, “software”

# Partitioning

- Do the “clumps” seen match our observations about the two main fields of application?
- Subject information not available in metadata
- Tag papers using journal titles (which are available)
  - Ex) “biology”, “cell”, “protein”, “DNA”
  - Ex) “computer vision”, “ACM”, “IEEE”, “software”
  - Gives us results for 54% of papers in the pruned network

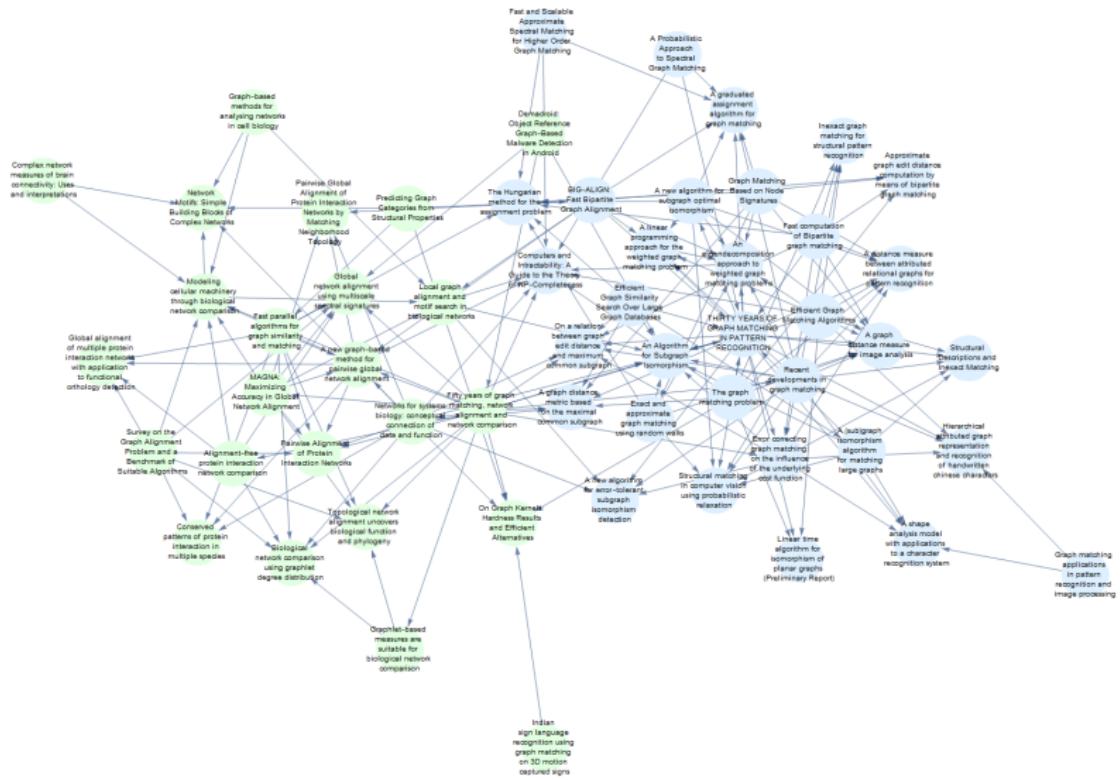
# Partitioning



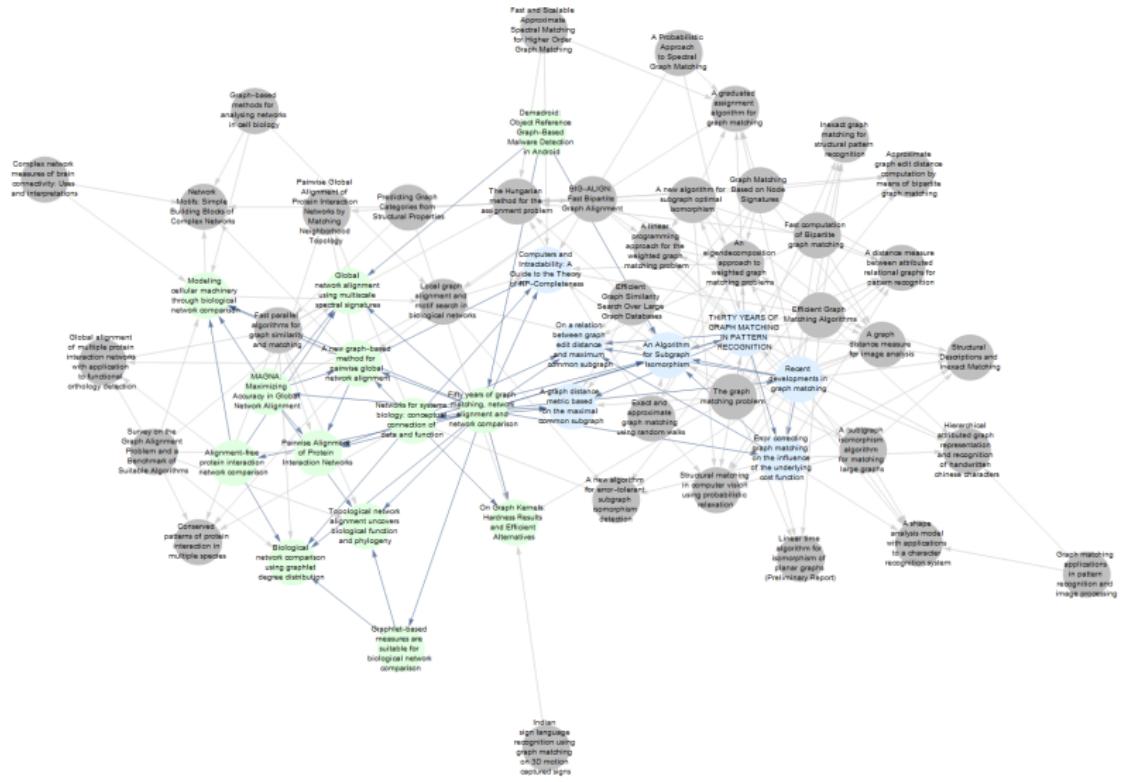
# Assortativity

	$G$	$G_p$
Outdegree	-0.0178	-0.0141
Publication year	0.0067	0.0041
Citation count	0.0006	0.0654
Reference count	0.0193	-0.0061
Tagged with any subject	0.1089	-0.0094
Subject	0.1837	0.0712
Subject is CS	0.2624	0.1529
Subject is biology	0.3354	0.1773
Subject is math	0.0732	0.0164
Subject is CS or biology	0.1500	0.0188
Subject is CS or math	0.2458	0.1256
Subject is biology or math	0.1713	0.0414

# Reading list subnetwork



# Reading list subnetwork



# Context

- Tabulated metadata (sort by year, title, etc.)
- Centrality ranks
- Neighbors within reading list
- Cocitations
- Global sense of structure via the partition

# Context

- Tabulated metadata (sort by year, title, etc.)
- Centrality ranks
- Neighbors within reading list
- Cocitations
- Global sense of structure via the partition

Using this context as we read allows us to draw connections and get a better sense of the shape of the field as a whole.

# Pattern recognition

- Graphs are a powerfully general data structure for the representation of objects and concepts

# Pattern recognition

- Graphs are a powerfully general data structure for the representation of objects and concepts
- Particularly useful in computer vision (invariant under positional transformations)

# Pattern recognition

- Graphs are a powerfully general data structure for the representation of objects and concepts
- Particularly useful in computer vision (invariant under positional transformations)
  - Optical character recognition
  - Biometric identification
  - 3D object recognition

# Pattern recognition

- Graphs are a powerfully general data structure for the representation of objects and concepts
- Particularly useful in computer vision (invariant under positional transformations)
  - Optical character recognition
  - Biometric identification
  - 3D object recognition
- Recognition/identification becomes a graph matching problem

# Graph matching terms

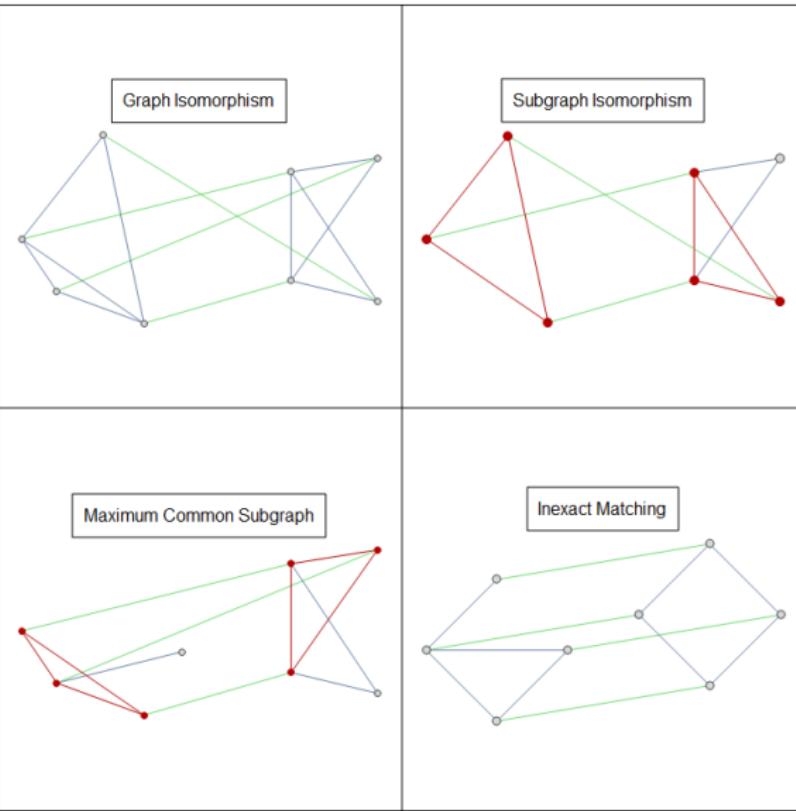
- Exact
- Inexact
- Error tolerant
- Optimal
- Approximate
- Alignment
- Elastic

# Exact

	Graph isomorphism	Subgraph isomorphism	Maximum common induced subgraph
$G_1$ and $G_2$ must have the same number of nodes	X		
Mapping must include all nodes of either $G_1$ or $G_2$	X	X	
Mapping must be edge-preserving	X	X	X
NP-complete	Unknown	X	X*

\*The associated decision problem of determining whether  $G_1$  and  $G_2$  have a common induced subgraph with at least  $k$  nodes is NP-complete, but the problem of finding the maximum common induced subgraph (as required for graph matching) is NP-hard

# Exact



# Comparison of formulations

	Edge preserving?	Result in?	Mapping seeking?	Optimal?	Complexity
Graph isomorphism	Yes	{0,1}	Yes	Yes	Likely between P and NP
Subgraph isomorphism	Yes	{0,1}	Yes	Yes	NP-complete
MCS computation	Yes	[0,1]	Yes	Yes	NP-hard
Edit distances (exact)	No	[0,1]	No	Yes	Generally exponential
Edit distances (approximate)	No	[0,1]	No	No	Generally polynomial
Other inexact formulations	No	[0,1]	Sometimes	No*	Generally polynomial

# Search space pruning

- Foundational algorithm for subgraph isomorphism is Ullmann 1976

# Search space pruning

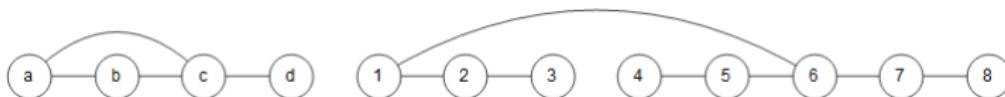
- Foundational algorithm for subgraph isomorphism is Ullmann 1976
- Rules out matching possibilities using two principles

# Search space pruning

- Foundational algorithm for subgraph isomorphism is Ullmann 1976
- Rules out matching possibilities using two principles
  - Degree comparison
  - Refinement

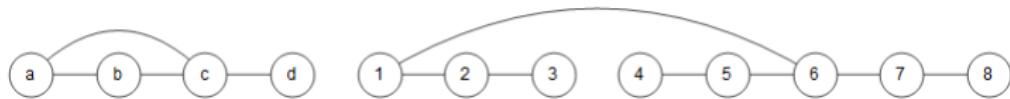
# Search space pruning

- Foundational algorithm for subgraph isomorphism is Ullmann 1976
- Rules out matching possibilities using two principles
  - Degree comparison
  - Refinement



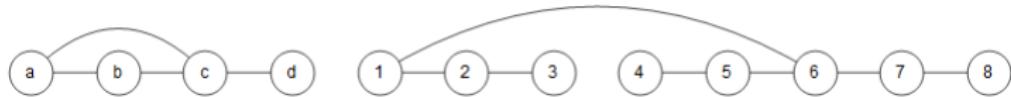
	1	2	3	4	5	6	7	8
a	1	1	0	0	1	1	1	0
b	1	1	0	0	1	1	1	0
c	0	0	0	0	0	1	0	0
d	1	1	1	1	1	1	1	1

# Refinement



	1	2	3	4	5	6	7	8
a	1	1	0	0	1	1	1	0
b	1	1	0	0	1	1	1	0
c	0	0	0	0	0	1	0	0
d	1	1	1	1	1	1	1	1

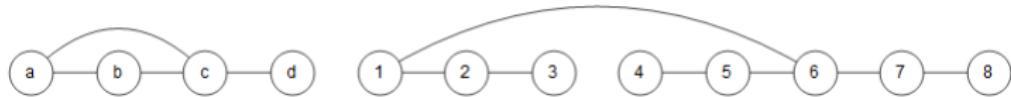
# Refinement



	1	2	3	4	5	6	7	8
a	1	1	0	0	1	1	1	0
b	1	1	0	0	1	1	1	0
c	0	0	0	0	0	1	0	0
d	1	1	1	1	1	1	1	1

	1	2	3	4	5	6	7	8
a	1	1	0	0	1	1	1	0
b	1	1	0	0	1	1	1	0
c	0	0	0	0	0	1	0	0
d	1	1	1	1	1	1	1	1

# Refinement

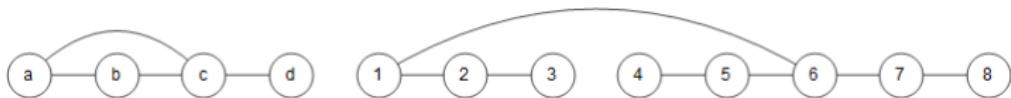


	1	2	3	4	5	6	7	8
a	1	1	0	0	1	1	1	0
b	1	1	0	0	1	1	1	0
c	0	0	0	0	0	1	0	0
d	1	1	1	1	1	1	1	1

	1	2	3	4	5	6	7	8
a	1	1	0	0	1	1	1	0
b	1	1	0	0	1	1	1	0
c	0	0	0	0	0	1	0	0
d	1	1	1	1	1	1	1	1

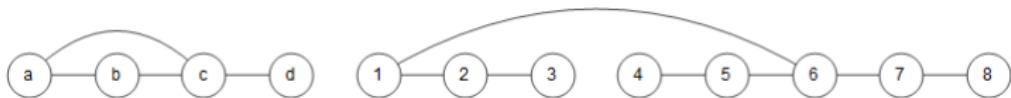
	1	2	3	4	5	6	7	8
a	1	0	0	0	1	0	1	0
b	1	0	0	0	1	0	1	0
c	0	0	0	0	0	1	0	0
d	1	0	0	0	1	0	1	0

# Backtracking



	1	2	3	4	5	6	7	8
a	1	0	0	0	1	0	1	0
b	0	0	0	0	1	0	1	0
c	0	0	0	0	0	1	0	0
d	0	0	0	0	1	0	1	0

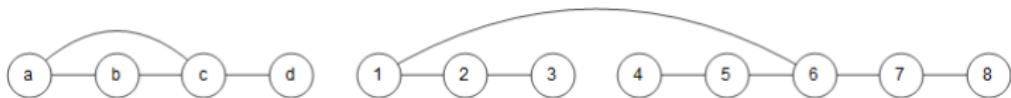
# Backtracking



	1	2	3	4	5	6	7	8
a	1	0	0	0	1	0	1	0
b	0	0	0	0	1	0	1	0
c	0	0	0	0	0	1	0	0
d	0	0	0	0	1	0	1	0

	1	2	3	4	5	6	7	8
a	0	0	0	0	1	0	1	0
b	1	0	0	0	0	0	1	0
c	0	0	0	0	0	1	0	0
d	1	0	0	0	0	0	1	0

# Backtracking



	1	2	3	4	5	6	7	8
a	1	0	0	0	1	0	1	0
b	0	0	0	0	1	0	1	0
c	0	0	0	0	0	1	0	0
d	0	0	0	0	1	0	1	0

	1	2	3	4	5	6	7	8
a	0	0	0	0	1	0	1	0
b	1	0	0	0	0	0	1	0
c	0	0	0	0	0	1	0	0
d	1	0	0	0	0	0	1	0

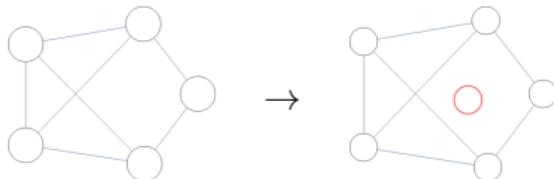
	1	2	3	4	5	6	7	8
a	0	0	0	0	0	0	1	0
b	1	0	0	0	1	0	0	0
c	0	0	0	0	0	1	0	0
d	1	0	0	0	1	0	0	0

# Edit distances

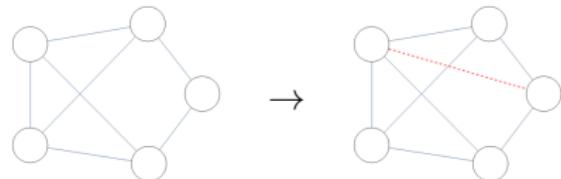
cat	→	cart	Insertion
cart	→	<i>dart</i>	Substitution
<i>dart</i>	→	art	Deletion
art	→	rat	Transposition

# Edit distances

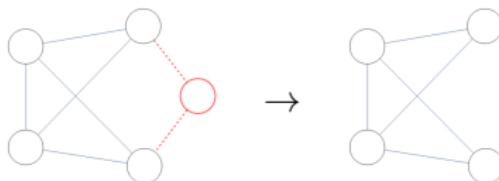
Node insertion



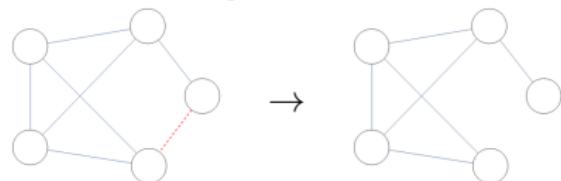
Edge insertion



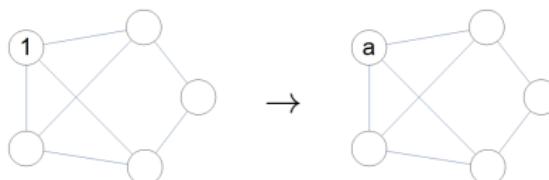
Node deletion



Edge deletion



Node substitution



# Approximate edit distances

- We can approximate the graph edit distance using the assignment problem

# Approximate edit distances

- We can approximate the graph edit distance using the assignment problem
- No longer an optimal method, but can be solved in  $O(n^3)$  time by the Hungarian method.

# Approximate edit distances

- We can approximate the graph edit distance using the assignment problem
- No longer an optimal method, but can be solved in  $O(n^3)$  time by the Hungarian method.

$$C = b \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 3 & 4 \\ 2 & 5 & 2 \end{pmatrix} \quad \Leftrightarrow \quad A = \{a, b, c\}, B = \{1, 2, 3\}$$

```
graph LR; a((a)) ---|3| 1((1)); a ---|2| 2((2)); a ---|2| 3((3)); b((b)) ---|1| 1; b ---|3| 2; b ---|4| 3; c((c)) ---|2| 3;
```

# Weighted graph matching problem

- Other suboptimal graph matching methods seek to solve the *weighted graph matching problem*.

# Weighted graph matching problem

- Other suboptimal graph matching methods seek to solve the *weighted graph matching problem*.
- Instead of indirectly optimizing the quality of a permutation assignment using a cost function, directly minimize  $\|A_G - PA_H P^T\|$  with respect to some norm and permutation matrix  $P$ , where  $A_G$  and  $A_H$  are the adjacency matrices of the weighted graphs  $G$  and  $H$ , respectively.

# Weighted graph matching problem

- Other suboptimal graph matching methods seek to solve the *weighted graph matching problem*.
- Instead of indirectly optimizing the quality of a permutation assignment using a cost function, directly minimize  $\|A_G - PA_H P^T\|$  with respect to some norm and permutation matrix  $P$ , where  $A_G$  and  $A_H$  are the adjacency matrices of the weighted graphs  $G$  and  $H$ , respectively.
- Avoids the heuristics inherent in cost function formulations, but we cannot incorporate external information (important for bio).

# Biology

- Biological network comparison is done primarily on protein-protein interaction (PPI) networks

# Biology

- Biological network comparison is done primarily on protein-protein interaction (PPI) networks
- Explosion of available PPI data parallels a previous surge of available DNA sequence information

# Biology

- Biological network comparison is done primarily on protein-protein interaction (PPI) networks
- Explosion of available PPI data parallels a previous surge of available DNA sequence information
- Networks are large, often incompletely explored, non-deterministic, and have real-world meaning

# Biology

- Biological network comparison is done primarily on protein-protein interaction (PPI) networks
- Explosion of available PPI data parallels a previous surge of available DNA sequence information
- Networks are large, often incompletely explored, non-deterministic, and have real-world meaning
- Overall goal is to organize data into models of basic biological processes

# Biology

- Biological network comparison is done primarily on protein-protein interaction (PPI) networks
- Explosion of available PPI data parallels a previous surge of available DNA sequence information
- Networks are large, often incompletely explored, non-deterministic, and have real-world meaning
- Overall goal is to organize data into models of basic biological processes
  - Determine function of proteins and groups of interactions
  - Investigate evolutionary relationships across species
  - Develop appropriate random graph models

# Biology

- Biological network comparison is done primarily on protein-protein interaction (PPI) networks
- Explosion of available PPI data parallels a previous surge of available DNA sequence information
- Networks are large, often incompletely explored, non-deterministic, and have real-world meaning
- Overall goal is to organize data into models of basic biological processes
  - Determine function of proteins and groups of interactions
  - Investigate evolutionary relationships across species
  - Develop appropriate random graph models
- Different strategies required compared to pattern recognition (small deterministic graphs)

# Strategies

- Count occurrences of very small subgraphs
  - Motifs and graphlets
  - Results in univariate statistics
  - Computational difficulty is in enumerating subgraphs, not matching them

# Strategies

- Count occurrences of very small subgraphs
  - Motifs and graphlets
  - Results in univariate statistics
  - Computational difficulty is in enumerating subgraphs, not matching them
- Local alignment
  - Find small highly conserved structures
  - Typically searches a deterministically constructed, not necessarily one-to-one mapping

# Strategies

- Count occurrences of very small subgraphs
  - Motifs and graphlets
  - Results in univariate statistics
  - Computational difficulty is in enumerating subgraphs, not matching them
- Local alignment
  - Find small highly conserved structures
  - Typically searches a deterministically constructed, not necessarily one-to-one mapping
- Global alignment
  - Find the best comprehensive superimposition of two input networks
  - Typically uses the assignment problem

# Subgraph counting

- Graphlets are small connected non-isomorphic induced subgraphs of a simple undirected network
- Can be used to generalize the degree distribution

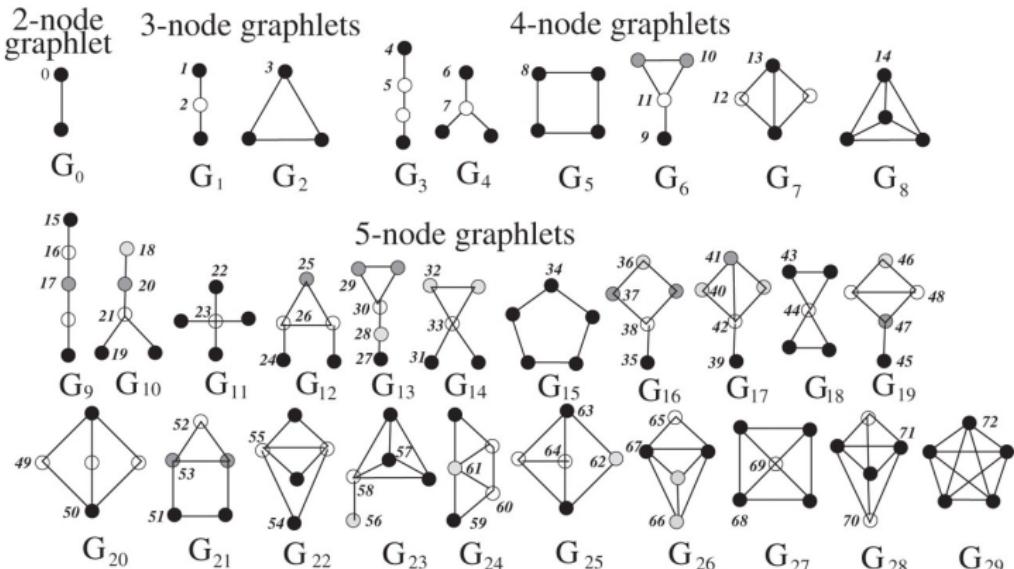
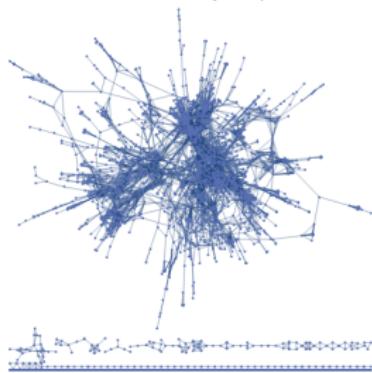


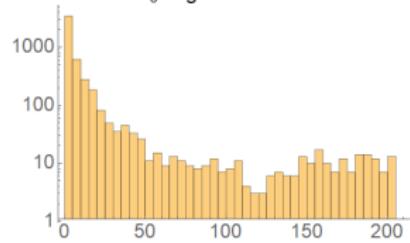
Figure sourced from "Biological Network Comparison Using Graphlet Degree Distribution" (Pržulj, 2007)

# Graphlets

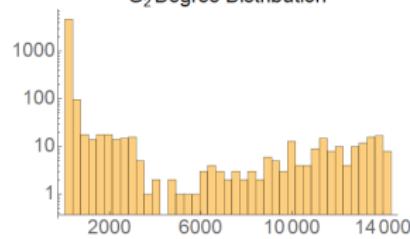
*Schizosaccharomyces pombe*



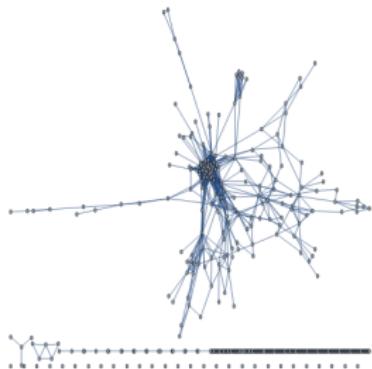
$G_0$  Degree Distribution



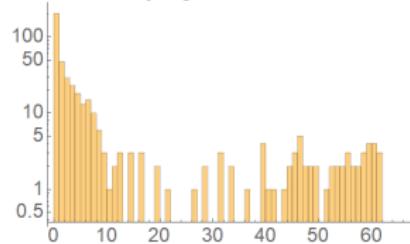
$G_2$  Degree Distribution



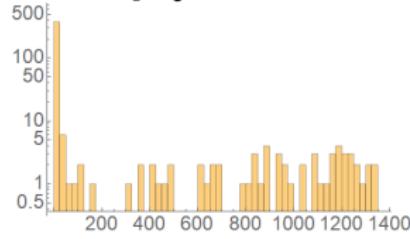
*Mycoplasma genitalium*



$G_0$  Degree Distribution

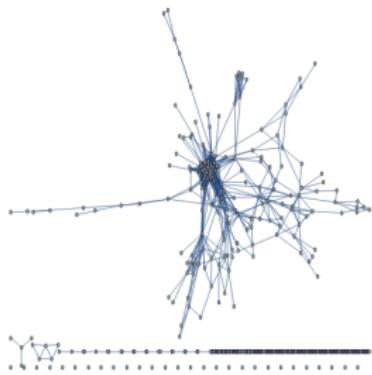


$G_2$  Degree Distribution

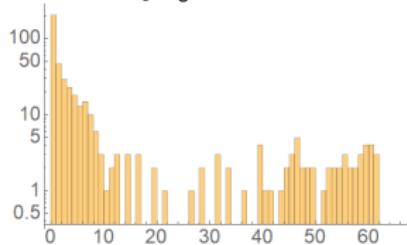


# Graphlets

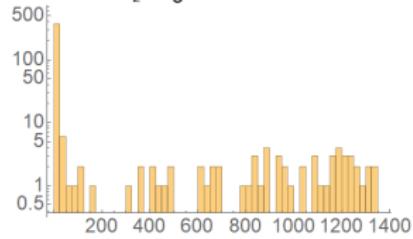
*Mycoplasma genitalium*



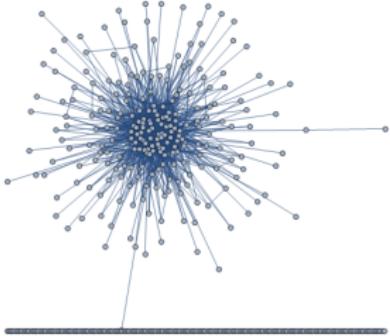
$G_0$  Degree Distribution



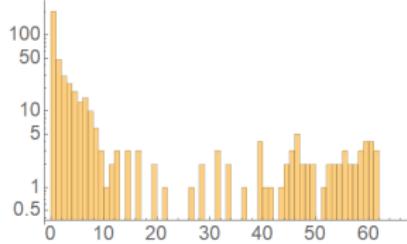
$G_2$  Degree Distribution



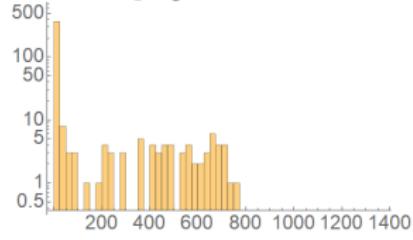
Random (Match Degree Distribution)



$G_0$  Degree Distribution



$G_2$  Degree Distribution

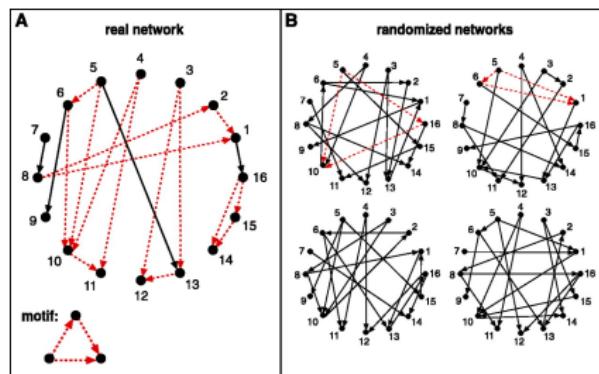
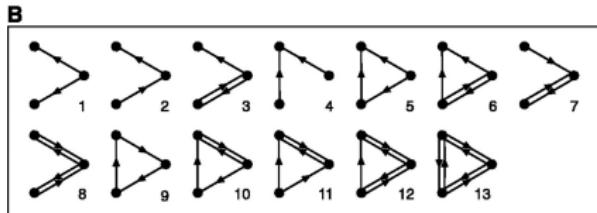
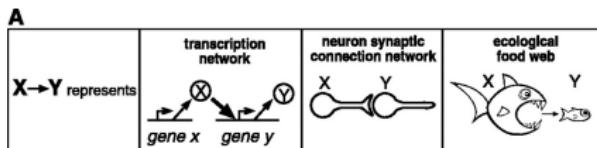


# Graphlets

	Vertices	Edges	Clustering	Maximum $G_0$ degree	Maximum $G_2$ degree
<i>Mycoplasma genitalium</i>	444	1860	0.758	66	1376
Random (match degree sequence)	444	1860	0.420	66	774
Random (match size)	444	1860	0.022	17	6
<i>S. pombe</i>	5100	30118	0.757	213	14592
Random (match degree sequence)	5100	30118	0.150	213	3606
Random (match size)	5100	30118	0.002	26	3

# Motifs

- Network motifs generalize sequence motifs
  - Find small patterns which are statistically overrepresented compared to a suitable null model
  - Can be directed, unlike graphlets, but do not have multiedges or self loops
  - Dependent on a good random model



Figures sourced from "Network Motifs: Simple Building Blocks of Complex Networks" (Milo et al., 2002)

# Analogy to sequence comparison

- The study of biological *networks* has been guided by the study of biological *sequences*

# Analogy to sequence comparison

- The study of biological *networks* has been guided by the study of biological *sequences*
- Sequence alignment is based on the assumptions that:

# Analogy to sequence comparison

- The study of biological *networks* has been guided by the study of biological *sequences*
- Sequence alignment is based on the assumptions that:
  - 1 Patterns which occur frequently are likely to have functional significance

# Analogy to sequence comparison

- The study of biological *networks* has been guided by the study of biological *sequences*
- Sequence alignment is based on the assumptions that:
  - 1 Patterns which occur frequently are likely to have functional significance
  - 2 Sequence regions conserved across multiple species are likely to have functional significance

# Analogy to sequence comparison

- The study of biological *networks* has been guided by the study of biological *sequences*
- Sequence alignment is based on the assumptions that:
  - 1 Patterns which occur frequently are likely to have functional significance
  - 2 Sequence regions conserved across multiple species are likely to have functional significance
  - 3 The degree to which sequences differ is related to their evolutionary distance

# Analogy to sequence comparison

- The study of biological *networks* has been guided by the study of biological *sequences*
- Sequence alignment is based on the assumptions that:
  - 1 Patterns which occur frequently are likely to have functional significance
  - 2 Sequence regions conserved across multiple species are likely to have functional significance
  - 3 The degree to which sequences differ is related to their evolutionary distance
- The study of motifs is based on assumption #1

# Alignment

- The development of biological network alignment closely parallels that of sequence alignment

# Alignment

- The development of biological network alignment closely parallels that of sequence alignment
- Problem

# Alignment

- The development of biological network alignment closely parallels that of sequence alignment
- Problem
  - Sequence alignment is  $O(mn)$
  - MCS problem is NP-hard

# Alignment

- The development of biological network alignment closely parallels that of sequence alignment
- Problem
  - Sequence alignment is  $O(mn)$
  - MCS problem is NP-hard
  - Alignment strategies must compromise

# Alignment

- The development of biological network alignment closely parallels that of sequence alignment
- Problem
  - Sequence alignment is  $O(mn)$
  - MCS problem is NP-hard
  - Alignment strategies must compromise
- Local alignment compromises:

# Alignment

- The development of biological network alignment closely parallels that of sequence alignment
- Problem
  - Sequence alignment is  $O(mn)$
  - MCS problem is NP-hard
  - Alignment strategies must compromise
- Local alignment compromises:
  - Search for a predetermined pattern
  - Construct alignment deterministically based on external information

# Alignment

- The development of biological network alignment closely parallels that of sequence alignment
- Problem
  - Sequence alignment is  $O(mn)$
  - MCS problem is NP-hard
  - Alignment strategies must compromise
- Local alignment compromises:
  - Search for a predetermined pattern
  - Construct alignment deterministically based on external information
- Global alignment compromise:

# Alignment

- The development of biological network alignment closely parallels that of sequence alignment
- Problem
  - Sequence alignment is  $O(mn)$
  - MCS problem is NP-hard
  - Alignment strategies must compromise
- Local alignment compromises:
  - Search for a predetermined pattern
  - Construct alignment deterministically based on external information
- Global alignment compromise:
  - Investigate assignment problem instead of maximum common induced subgraph problem

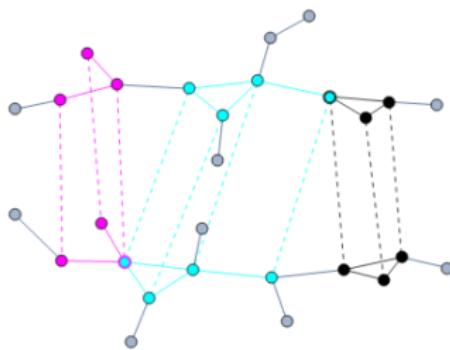
# Alignment

- Local

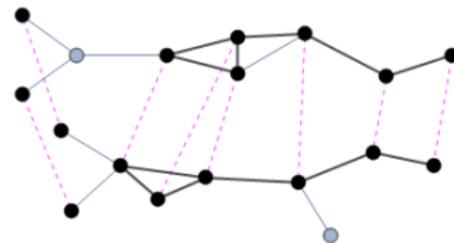
- Search for local regions of isomorphism
- Mappings do not need to be mutually consistent

- Global

- Single mapping across all parts of the input
- Large but suboptimally conserved mapped subnetworks



Local alignment



Global alignment

# Local alignment

- PathBLAST (2004)
- NetworkBLAST (2008)
- Graemlin (2006)
- MAWISH (2006)

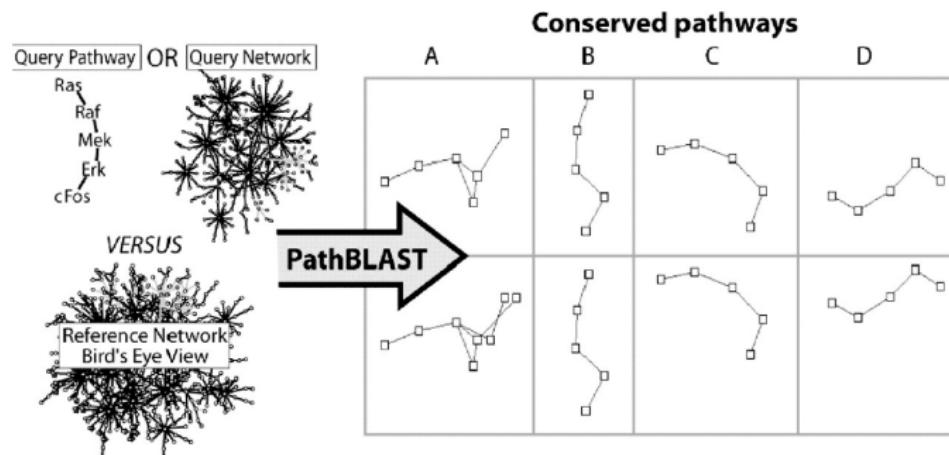


Figure sourced from "PathBLAST: A Tool For Alignment of Protein Interaction Networks" (Kelley et al., 2004)

# Local alignment

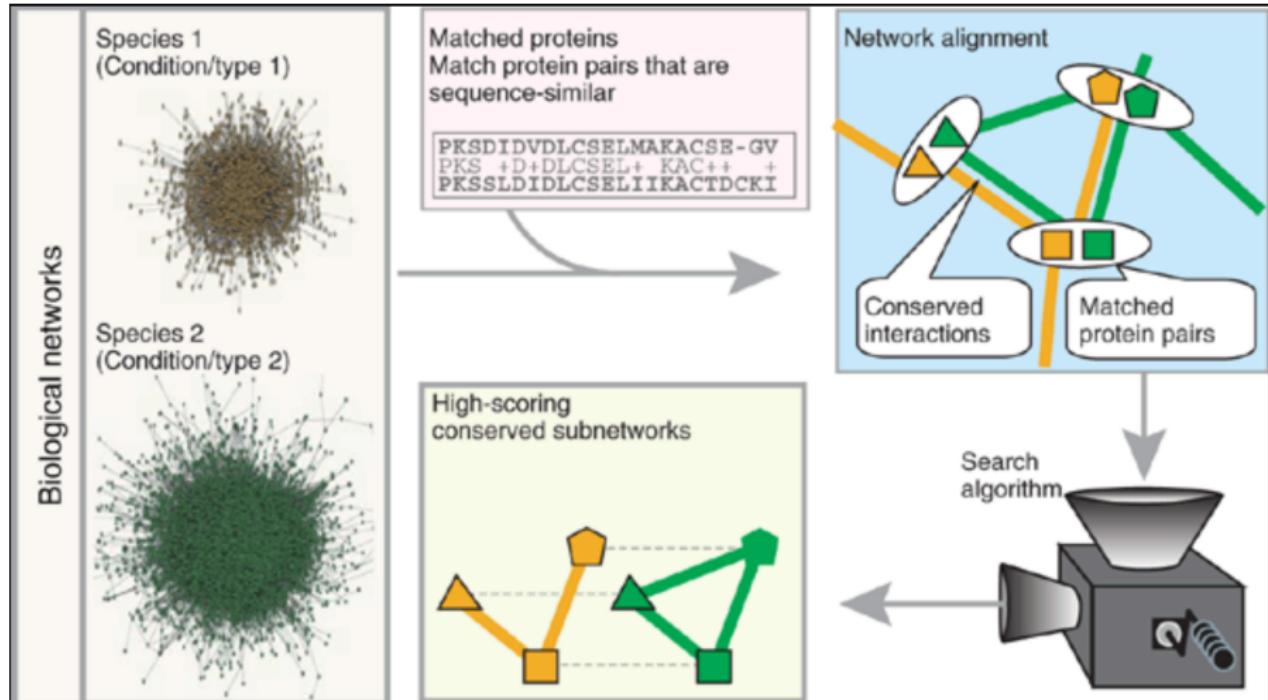


Figure sourced from "Modeling cellular machinery through biological network comparison" (Sharan and Ideker, 2006)

# Global alignment

- Formulate as the assignment problem

# Global alignment

- Formulate as the assignment problem
- Two steps
  - Cost matrix
  - Construct mapping

# Global alignment

- Formulate as the assignment problem
- Two steps
  - Cost matrix
  - Construct mapping
- How does biology's use of the assignment problem differ from pattern recognition?

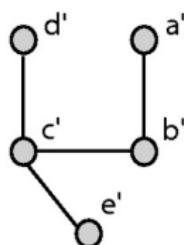
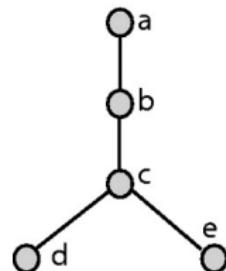
# Global alignment

- Formulate as the assignment problem
- Two steps
  - Cost matrix
  - Construct mapping
- How does biology's use of the assignment problem differ from pattern recognition?
  - External node info as well as topological
  - Various measures of alignment quality
  - Neighborhood-based mapping construction
  - Far more common

# Global alignment algorithms discussed

- Biology
  - IsoRank (2007)
  - Natalie (2009)
  - GRAAL (2010)
  - PINALOG (2012)
  - GHOST (2012)
  - SPINAL (2013)
  - NETAL (2013)
  - MAGNA (2014)
- Non-biology
  - Node signatures (2009)
  - GED approximation (2009)
  - Modified GED approximation (2014)

# IsoRank intuition



R

	a'	b'	c'	d'	e'
a	0.0312		0.0937		
b		0.1250		0.0625	0.0625
c	0.0937		0.2812		
d		0.0625		0.0312	0.0312
e		0.0625		0.0312	0.0312

$$R_{aa'} = \frac{1}{4} R_{bb'}$$

$$R_{bb'} = \frac{1}{3} R_{ac'} + \frac{1}{3} R_{a'c} + R_{aa'} + \frac{1}{9} R_{cc'}$$

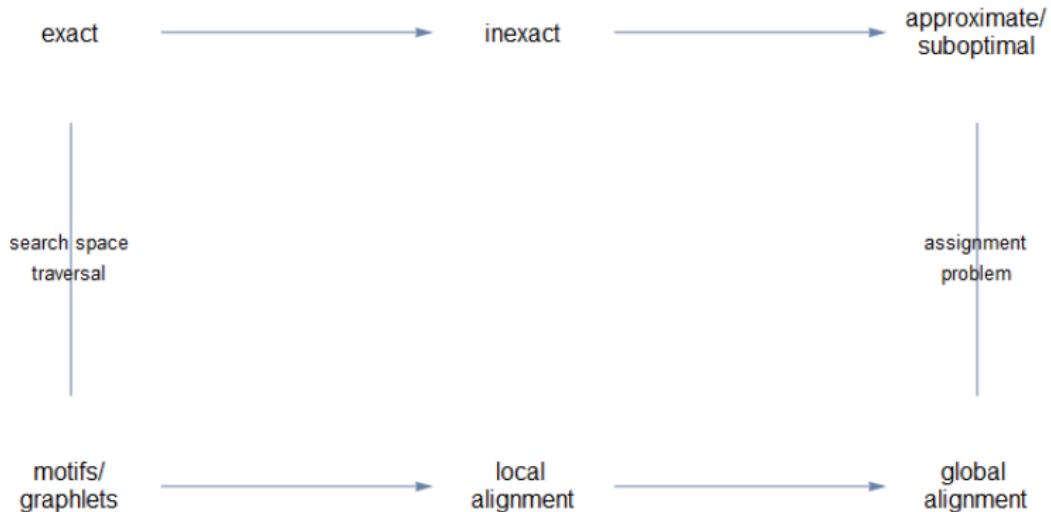
$$R_{dd'} = \frac{1}{9} R_{cc'}$$

$$R_{cc'} = \frac{1}{4} R_{bb'} + \frac{1}{2} R_{be'} + \frac{1}{2} R_{bd'} + \frac{1}{2} R_{eb'} + \frac{1}{2} R_{db'} + R_{ee'} + R_{ed'} + R_{de'} + R_{dd'}$$

Figure sourced from "Global alignment of multiple protein interaction networks with application to functional orthology detection" (Singh et al., 2008)

$$R_{ij} = \sum_{u \in N(i)} \sum_{v \in N(j)} \frac{R_{uv}}{|N(u)||N(v)|}, \quad i \in V_1, j \in V_2,$$

# Conclusion



# Conclusion

- Networks being investigated inform approach used

# Conclusion

- Networks being investigated inform approach used
- Potential cross-applications
  - Graphlets for pattern recognition/database search
  - Global alignment techniques for assignment problem
  - Biological strategies for social networks

# Conclusion

- Networks being investigated inform approach used
- Potential cross-applications
  - Graphlets for pattern recognition/database search
  - Global alignment techniques for assignment problem
  - Biological strategies for social networks
- Utility of citation network method

# Questions?