

A computationally driven comparative survey of network alignment, graph matching, and  
network comparison in pattern recognition and systems biology

Marissa Graham

A thesis submitted to the faculty of  
Brigham Young University  
in partial fulfillment of the requirements for the degree of  
Master of Science

Emily Evans, Chair  
Benjamin Webb  
Christopher Grant

Department of Mathematics  
Brigham Young University

Copyright © 2018 Marissa Graham

All Rights Reserved

# CONTENTS

Contents	ii
List of Tables	iii
List of Figures	v
Bibliography	12
Index	17

## LIST OF TABLES

1	Comparing statistics for our dataset to other networks. . . . .	1
2	Highest centrality papers for the entire pruned network. . . . .	2
3	Highest centrality papers for Group 1 (biology dominated) in our partition of the pruned network. . . . .	3
4	Highest centrality papers for Group 2 (computer science dominated) in our partition of the pruned network. . . . .	4
5	Number of vertices tagged as computer science, biology, math, or some combination of these in $G$ , $G_p$ , and the two halves of the partition $G_p^{(1)}$ and $G_p^{(2)}$ . . . . .	5
6	Assortativity of the full and pruned citation networks with respect to various network properties. . . . .	5
7	A summary of exact graph matching problem formulations. . . . .	6
8	Summary of the distinctions between exact and inexact graph matching styles. . . . .	6
9	A summary of the distinction between local and global network statistics. . . . .	7
10	Statistics for PPI networks of two small organisms and two comparable random graphs for each. The “clustering” value is the global clustering coefficient [43], which measures the fraction of connected triplets in the network which are closed. Edge density is the fraction of edges compared to the total number of possible edges, i.e., $m/n^2$ . . . . .	8
11	Broad summary of alignment algorithms discussed in this section. The distinctions between the various topological similarity scores used are discussed in each algorithm’s individual section. . . . .	9
12	Number of vertices tagged as computer science, biology, math, or some combination of these in the reading list subnetwork $G_R$ , and its intersections $G_R^{(1)}$ and $G_R^{(2)}$ with the two halves of the partition $G_p^{(1)}$ and $G_p^{(2)}$ . See Table 5. . . . .	10

13	Python packages used for the project. All but those marked with a $\diamond$ are found in either the standard library or available in Anaconda for Python 3.6 on 64-bit Windows in mid-2018, and the remainder can be installed via pip.	10
14	Keywords used to tag journal names as various subjects. . . . .	11

## LIST OF FIGURES

1	Edit operations for strings. . . . .	5
2	A summary of the distinction between univariate and bivariate measures. . .	7

	$G$	$G_p$	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
Fraction in giant component	0.960	1.000	0.784	0.797	0.900	0.999

Table 1: Comparing statistics for our dataset to other networks.

Title	Indegree	Outdegree	Betweenness	Closeness	HITS Auth.	HITS Hub
◊Thirty Years of Graph Matching in Pattern Recognition [14]	20*	109*	1	2		1
†Fifty years of graph matching, network alignment and network comparison [21]	6	71*	2	1		3
†Networks for systems biology: conceptual connection of data and function [20]	2	102*	3	3		2
◊An Algorithm for Subgraph Isomorphism [61]	20*	4	7	4	1	
†Modeling cellular machinery through biological network comparison [58]	9	41*	8			
◊Computers and Intractability: A Guide to the Theory of NP-Completeness [26]	16*	0	4	5		
◊The graph matching problem [38]	2	55*	5	6		7
†A new graph-based method for pairwise global network alignment [30]	9	13		8		
†On Graph Kernels: Hardness Results and Efficient Alternatives [25]	11	10	6			
◊Error correcting graph matching: on the influence of the underlying cost function [8]	10	16		7	7	8
◊A graduated assignment algorithm for graph matching [23]	18*	0			5	
◊The Hungarian method for the assignment problem [35]	17*	0				
◊An eigendecomposition approach to weighted graph matching problems [62]	15*	5			6	
◊Recent developments in graph matching [6]	1	51*				4
†MAGNA: Maximizing Accuracy in Global Network Alignment [54]	5	35*				
◊A distance measure between attributed relational graphs for pattern recognition [53]	14*	0			3	
†Pairwise Global Alignment of Protein Interaction Networks by Matching Neighborhood Topology [60]	13*	0				
†Topological network alignment uncovers biological function and phylogeny [10]	12*	0				
A graph distance metric based on the maximal common subgraph [34]	10	0		10	4	
◊Efficient Graph Matching Algorithms [40]	0	43*				5
Local graph alignment and motif search in biological networks [5]	8	10	10			
†Global alignment of multiple protein interaction networks with application to functional orthology detection [59]	11*	0				
On a relation between graph edit distance and maximum common subgraph [7]	11	0			2	
◊Graph matching applications in pattern recognition and image processing [13]	0	40*				6
◊Fast and Scalable Approximate Spectral Matching for Higher Order Graph Matching [45]	0	41*	9			
◊Structural matching in computer vision using probabilistic relaxation [12]	9	0			10	
◊A new algorithm for subgraph optimal isomorphism [19]	2	21				9
BIG-ALIGN: Fast Bipartite Graph Alignment [32]	2	21		9		
◊A graph distance measure for image analysis [22]	8	0			8	
A new algorithm for error-tolerant subgraph isomorphism detection [41]	8	0			9	
◊A (sub)graph isomorphism algorithm for matching large graphs [15]	3	16				10

†Also top for Group 1 (biology dominated); ◊Also top for Group 2 (computer science dominated)

Table 2: Highest centrality papers for the entire pruned network.

Title	Indegree	Outdegree	Betweenness	Closeness	HITS Auth.	HITS Hub
◊Networks for systems biology: conceptual connection of data and function [20]	2	90*	1	2		1
◊Fifty years of graph matching, network alignment and network comparison [21]	4	56*	2	1		2
◊Modeling cellular machinery through biological network comparison [58]	9	40*	4	3	10	9
◊MAGNA: Maximizing Accuracy in Global Network Alignment [54]	5	35*	7	6		3
◊On Graph Kernels: Hardness Results and Efficient Alternatives [25]	10*	9	3	8		
Biological network comparison using graphlet degree distribution [48]	11*	0		7	4	7
◊A new graph-based method for pairwise global network alignment [30]	8	12	9	4	6	
Network Motifs: Simple Building Blocks of Complex Networks [42]	11*	0		9	8	
◊Pairwise Global Alignment of Protein Interaction Networks by Matching Neighborhood Topology [60]	12*	0			3	
◊Topological network alignment uncovers biological function and phylogeny [34]	12*	0			2	
NETAL: a new graph-based method for global alignment of protein-protein interaction networks [44]	6	26*				5
Collective dynamics of “small-world” networks [64]	10*	0		10	5	
Global network alignment using multiscale spectral signatures [46]	11*	0			9	
◊Global alignment of multiple protein interaction networks with application to functional orthology detection [59]	10*	0				
Conserved patterns of protein interaction in multiple species [57]	10*	0			7	
Pairwise Alignment of Protein Interaction Networks [33]	10*	0			1	
Alignment-free protein interaction network comparison [3]	2	22	6	5		
Graphlet-based measures are suitable for biological network comparison [27]	1	30*				8
Survey on the Graph Alignment Problem and a Benchmark of Suitable Algorithms [17]	0	26				4
Predicting Graph Categories from Structural Properties [11]	0	30*	5			
Fast parallel algorithms for graph similarity and matching [31]	1	23				6
Complex network measures of brain connectivity: Uses and interpretations [52]	0	28*	8			
Graph-based methods for analysing networks in cell biology [1]	0	30*				10
Demadroid: Object Reference Graph-Based Malware Detection in Android [63]	0	25	10			
Early Estimation Model for 3D-Discrete Indian Sign Language Recognition Using Graph Matching [37]	0	29*				
Indian sign language recognition using graph matching on 3D motion captured signs [36]	0	29*				

◊Also a top-centrality paper for the entire network

Table 3: Highest centrality papers for Group 1 (biology dominated) in our partition of the pruned network.



Title	Indegree	Outdegree	Betweenness	Closeness	HITS Auth.	HITS Hub
◊Thirty Years of Graph Matching in Pattern Recognition [14]	17*	107*	1	1		1
◊An Algorithm for Subgraph Isomorphism [61]	15*	2	10	5	2	
◊A graduated assignment algorithm for graph matching [23]	18*	0	7	4	3	
◊An eigendecomposition approach to weighted graph matching problems [62]	15*	5		2	4	
◊The graph matching problem [38]	2	36*	3	3		8
◊A distance measure between attributed relational graphs for pattern recognition [53]	13*	0		7	1	
◊Recent developments in graph matching [6]	0	50*	8			2
◊Error correcting graph matching: on the influence of the underlying cost function [8]	9*	16		8		6
◊Fast and Scalable Approximate Spectral Matching for Higher Order Graph Matching [45]	0	41*	2			
◊Efficient Graph Matching Algorithms [40]	0	42*	5			4
◊Computers and Intractability: A Guide to the Theory of NP-Completeness [26]	11*	0	6			
◊The Hungarian method for the assignment problem [35]	14*	0				
◊Graph matching applications in pattern recognition and image processing [13]	0	40*				3
Efficient Graph Similarity Search Over Large Graph Databases [66]	0	28*	4	6		
A linear programming approach for the weighted graph matching problem [4]	8	8		9	9	
◊Structural matching in computer vision using probabilistic relaxation [12]	9*	0			5	
◊A graph distance measure for image analysis [22]	8	0			6	
Inexact graph matching for structural pattern recognition [9]	10*	0				
◊A new algorithm for subgraph optimal isomorphism [19]	2	21				5
Approximate graph edit distance computation by means of bipartite graph matching [50]	9	0				
Linear time algorithm for isomorphism of planar graphs (Preliminary Report) [28]	9	0				
Structural Descriptions and Inexact Matching [56]	9	0			7	
◊A (sub)graph isomorphism algorithm for matching large graphs [15]	3	16				7
A Probabilistic Approach to Spectral Graph Matching [18]	0	25*	9	10		
Hierarchical attributed graph representation and recognition of handwritten chinese characters [39]	6	0			8	
Exact and approximate graph matching using random walks [24]	1	14				9
A shape analysis model with applications to a character recognition system [51]	5	0			10	
Fast computation of Bipartite graph matching [55]	1	23*				
Graph Matching Based on Node Signatures [29]	0	17				10
Unsupervised Domain Adaptation Using Regularized Hyper-Graph Matching [16]	0	22*				

◊Also a top-centrality paper for the entire network

Table 4: Highest centrality papers for Group 2 (computer science dominated) in our partition of the pruned network.

	$G$	$G_p$	$G_p^{(1)}$	$G_p^{(2)}$
Total vertices	5793	1062	531	531
Untagged	1922	502	311	191
Tagged	3871	560	220	340
CS	2533	405	93	312
Biology	984	122	108	14
Math	787	97	44	53
Both CS and biology	108	13	9	4
Both CS and math	305	49	15	34
Both biology and math	24	3	2	1
All three	4	1	1	0

Table 5: Number of vertices tagged as computer science, biology, math, or some combination of these in  $G$ ,  $G_p$ , and the two halves of the partition  $G_p^{(1)}$  and  $G_p^{(2)}$ .

	$G$	$G_p$
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

Table 6: Assortativity of the full and pruned citation networks with respect to various network properties.

### 0.0.1 Exact and inexact matching.

cat → cart	Insertion
cart → dart	Substitution
dart → art	Deletion
art → rat	Transposition

Figure 1: Edit operations for strings.

	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 [65].

Table 7: A summary of exact graph matching problem 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

Table 8: Summary of the distinctions between exact and inexact graph matching styles.

\*The Hungarian algorithm can be used to find an optimal assignment in  $O(n^3)$  time based on a given cost function, but the assignment problem minimizes a cost function which is only an approximation of the true matching cost.

## 0.0.2 Graph edit distance.

### Network comparison with univariate measures

Network  $\rightarrow$  Something in  $\mathbb{R}^n$   $\left. \vphantom{\begin{matrix} \text{Network} \\ \text{Network} \end{matrix}} \right\} \rightarrow$  Similarity score derived from a metric  
 Network  $\rightarrow$  Something in  $\mathbb{R}^n$  or aggregation measure on  $\mathbb{R}^n$

Examples: Any metric on  $\mathbb{R}^n$  applied to number of vertices and edges, mean degree, diameter, connectivity, degree distribution, centrality distributions, local graphlet counts, graphlet degree distributions, etc.

### Network comparison with bivariate measures

Network  $\left. \vphantom{\begin{matrix} \text{Network} \\ \text{Network} \end{matrix}} \right\} \rightarrow$  Mapping process  $\rightarrow$  Similarity measure derived from  
 Network of some sort the mapping in some way

Examples: Graph edit distance; node correctness, edge correctness, induced conserved structure [46], or symmetric substructure score [54] with respect to an alignment; MCS-related metrics (i.e., [10])

Figure 2: A summary of the distinction between univariate and bivariate measures.

Type	Description	Examples
Global	Single value for an entire network	Mean degree, maximum degree, diameter, edge density, assortativity, global clustering coefficient
Local	Value at each vertex in a network	Indegree, outdegree, graphlet degrees, betweenness centrality, closeness centrality, HITS centralities, local clustering coefficient

Table 9: A summary of the distinction between local and global network statistics.

	Vertices	Edges	Edge density	Clustering	Maximum $G_0$ degree	Maximum $G_2$ degree
<i>Mycoplasma genitalium</i>	444	1860	0.94%	0.758	66	1376
Match degree sequence	444	1860	0.94%	0.420	66	774
Match size and density	444	1860	0.94%	0.022	17	6
<i>S. pombe</i>	5100	30118	0.11%	0.757	213	14592
Match degree sequence	5100	30118	0.11%	0.150	213	3606
Match size and density	5100	30118	0.11%	0.002	26	3

Table 10: Statistics for PPI networks of two small organisms and two comparable random graphs for each. The “clustering” value is the global clustering coefficient [43], which measures the fraction of connected triplets in the network which are closed. Edge density is the fraction of edges compared to the total number of possible edges, i.e.,  $m/n^2$ .

Biology	Year	Similarity Scoring	Alignment Construction
IsoRank [60]	2007	Convex combination of external information and eigenvalue problem-based topological node similarities	Maximum-weight bipartite matching OR Repeated greedy pairing of highest scores
Natalie [30]	2009	Convex combination of external info-based node mapping scores and topology-based edge mapping scores	Cast as an integer linear programming problem and use Lagrangian relaxation
GRAAL [34]	2010	Convex combination of graphlet signatures and local density	Greedy neighborhood alignment around highest-scoring pairs
PINALOG [47]	2012	Only sequence and functional similarity of proteins initially, but includes topological similarity for extension mapping	Detect communities, pair similar proteins from communities, extend the mapping to their neighbors
GHOST [46]	2012	Eigenvalue distributions of appropriately normalized neighborhood Laplacians	Seed-and-extend with approximate solutions to the QAP, then local search step
SPINAL [2]	2013	Convex combination of sequence similarity and neighbor matching-based topological similarity	Seed-and-extend with local swaps
NETAL [44]	2013	Update an initial scoring based on the fraction of common neighbors between matched pairs in its corresponding greedy alignment	Repeated greedy pairing of highest scores, while updating expected number of conserved interactions
MAGNA [54]	2014	Any	Improve a population of existing alignments with crossover and a fitness function
Node fingerprinting [49]	2014	Minimize degree differences and reward adjacency to already-matched pairs	Progressively add high-scoring pairings to an alignment and update scores
Non-biology	Year	Similarity Scoring	Alignment Construction
Node signatures [29]	2009	Vertex degree and incident edge weights	Hungarian method
Graph edit distance approximation [50]	2009	Edit costs (vertex insertions, substitutions, deletions)	Generalized (non-square) Munkres' algorithm
Modified GED approximation [55]	2014	Modification of edit costs when edit distance is a proper distance function	Generalized (non-square) Munkres' algorithm

Table 11: Broad summary of alignment algorithms discussed in this section. The distinctions between the various topological similarity scores used are discussed in each algorithm's individual section.

	$G_R$	$G_R^{(1)}$	$G_R^{(2)}$
Total vertices	61	27	34
Untagged	30	8	22
Tagged	31	19	12
CS	24	2	22
Biology	6	6	0
Math	3	1	2
Both CS and biology	1	1	0
Both CS and math	2	0	2
Both biology and math	0	0	0
All three	0	0	0

Table 12: Number of vertices tagged as computer science, biology, math, or some combination of these in the reading list subnetwork  $G_R$ , and its intersections  $G_R^{(1)}$  and  $G_R^{(2)}$  with the two halves of the partition  $G_p^{(1)}$  and  $G_p^{(2)}$ . See Table 5.

Usage	Package Name(s)
Mathematical Computation	NumPy NetworkX
Figure creation*	Matplotlib
File I/O Handling	csv glob re
API Request Handling	urllib Requests time
Interfacing with Google Sheets	gsread $\diamond$ oauth2client $\diamond$
The <code>defaultdict</code> datatype	collections
Interfacing my own modules with Jupyter notebooks	importlib $\diamond$

Table 13: Python packages used for the project. All but those marked with a  $\diamond$  are found in either the standard library or available in Anaconda for Python 3.6 on 64-bit Windows in mid-2018, and the remainder can be installed via pip.

\*Only Figure ?? was created in Python. The remainder were made in Mathematica.

Computer Science	Biology	Mathematics
ACM	Biochem-	Algebra
Algorithm	Biocomputing	Algorithm
Artificial Intelligence	Bioengineering	Chaos
CIVR	Bioinformatic	Combinatori-
Computational Intelligence	Biological	Fixed Point
Computational Linguistics	Biology	Fractal
Computer	Biomedic-	Functional Analysis
Computer Graphics	Biosystem	Geometr-
Computer Science	Biotechnology	Graph
Computer Vision	Brain	Kernel
Data	Cancer	Linear Regression
Data Mining	Cardiology	Markov
Document Analysis	Cell	Mathemati-
Electrical Engineering	Disease	Multivariate
Graphics	DNA	Network
IEEE	Drug	Optimization
Image Analysis	Endocrinology	Permutation Group
Image Processing	Epidemiology	Probability
Intelligent System	Genetic	Riemann Surface
Internet	Genome	SIAM
ITiCSE	Genomic	Statistic-
Language Processing	Medical	Topology
Learning	Medicinal	Wavelet
Machine Learning	Medicine	
Machine Vision	Metabolic	
Malware	Microbiology	
Neural Network	Molecular	
Pattern Recognition	Neuro-	
Robotic	Neurobiological	
Scientific Computing	Pathology	
SIAM	Pathogen	
Signal Processing	Pharma-	
Software	Plant	
World Wide Web	Protein	
	Proteom-	
	Psych-	
	Psychology	
	Virology	
	Virus	

Table 14: Keywords used to tag journal names as various subjects.

\*Note: Both a term and its plural are considered a match, and hyphens indicate a word with several ending variations which were all considered to be associated with the tag. While the search process was case sensitive in order to avoid false positives for short words like “ACM”, case-insensitive duplicate words have been excluded from the table. The words “algorithm” and “SIAM” are considered to be both computer science and mathematics.



## BIBLIOGRAPHY

- [1] T. Aittokallio. Graph-based methods for analysing networks in cell biology. *Briefings in Bioinformatics*, 7(3):243–255, May 2006.
- [2] Ahmet E Aladağ and Cesim Erten. Spinal: scalable protein interaction network alignment. *Bioinformatics*, 29(7):917–924, 2013.
- [3] Waqar Ali, Tiago Rito, Gesine Reinert, Fengzhu Sun, and Charlotte M. Deane. Alignment-free protein interaction network comparison. *Bioinformatics*, 30(17):i430–i437, August 2014.
- [4] H.A. Almohamad and S.O. Duffuaa. A linear programming approach for the weighted graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5):522–525, May 1993.
- [5] J. Berg and M. Lassig. Local graph alignment and motif search in biological networks. *Proceedings of the National Academy of Sciences*, 101(41):14689–14694, September 2004.
- [6] H. Bunke. Recent developments in graph matching. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*. IEEE Comput. Soc.
- [7] H. Bunke. On a relation between graph edit distance and maximum common subgraph. *Pattern Recognition Letters*, 18(8):689–694, August 1997.
- [8] H. Bunke. Error correcting graph matching: on the influence of the underlying cost function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):917–922, 1999.
- [9] H Bunke and G Allermann. Inexact graph matching for structural pattern recognition. *Pattern Recognition Letters*, 1(4):245–253, May 1983.
- [10] Horst Bunke and Kim Shearer. A graph distance metric based on the maximal common subgraph. *Pattern Recognition Letters*, 19(3-4):255–259, March 1998.
- [11] James P Canning, Emma E Ingram, Sammantha Nowak-Wolff, Adriana M Ortiz, Nesreen K Ahmed, Ryan A Rossi, Karl RB Schmitt, and Sucheta Soundarajan. Predicting graph categories from structural properties. *arXiv preprint arXiv:1805.02682*, 2018.
- [12] W.J. Christmas, J. Kittler, and M. Petrou. Structural matching in computer vision using probabilistic relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):749–764, 1995.
- [13] D. Conte, P. Foggia, C. Sansone, and M. Vento. Graph matching applications in pattern recognition and image processing. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*. IEEE.
- [14] D. CONTE, P. FOGGIA, C. SANSONE, and M. VENTO. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(03):265–298, May 2004.

- [15] L.P. Cordella, P. Foggia, C. Sansone, and M. Vento. A (sub)graph isomorphism algorithm for matching large graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1367–1372, October 2004.
- [16] Debasmith Das and CS Lee. Unsupervised domain adaptation using regularized hyper-graph matching. *arXiv preprint arXiv:1805.08874*, 2018.
- [17] Christoph Döpmann. Survey on the graph alignment problem and a benchmark of suitable algorithms. *Institut für Informatik*, 2013.
- [18] Amir Egozi, Yosi Keller, and Hugo Guterman. A probabilistic approach to spectral graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):18–27, January 2013.
- [19] Yasser El-Sonbaty and M.A. Ismail. A new algorithm for subgraph optimal isomorphism. *Pattern Recognition*, 31(2):205–218, February 1998.
- [20] F. Emmert-Streib and M. Dehmer. Networks for systems biology: conceptual connection of data and function. *IET Systems Biology*, 5(3):185–207, May 2011.
- [21] Frank Emmert-Streib, Matthias Dehmer, and Yongtang Shi. Fifty years of graph matching, network alignment and network comparison. *Information Sciences*, 346-347:180–197, June 2016.
- [22] M. A. Eshera and King-Sun Fu. A graph distance measure for image analysis. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-14(3):398–408, May 1984.
- [23] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):377–388, April 1996.
- [24] M. Gori, M. Maggini, and L. Sarti. Exact and approximate graph matching using random walks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(7):1100–1111, July 2005.
- [25] Thomas Gärtner, Peter Flach, and Stefan Wrobel. On graph kernels: Hardness results and efficient alternatives. In *Learning Theory and Kernel Machines*, pages 129–143. Springer Berlin Heidelberg, 2003.
- [26] Juris Hartmanis. Computers and intractability: A guide to the theory of np-completeness (michael r. garey and david s. johnson). *SIAM Review*, 24(1):90–91, January 1982.
- [27] W. Hayes, K. Sun, and N. Przulj. Graphlet-based measures are suitable for biological network comparison. *Bioinformatics*, 29(4):483–491, January 2013.
- [28] J. E. Hopcroft and J. K. Wong. Linear time algorithm for isomorphism of planar graphs (preliminary report). In *Proceedings of the sixth annual ACM symposium on Theory of computing - STOC '74*. ACM Press, 1974.

- [29] Salim Jouili and Salvatore Tabbone. Graph matching based on node signatures. In *Graph-Based Representations in Pattern Recognition*, pages 154–163. Springer Berlin Heidelberg, 2009.
- [30] Gunnar W Klau. A new graph-based method for pairwise global network alignment. *BMC Bioinformatics*, 10(Suppl 1):S59, 2009.
- [31] Giorgos Kollias, Madan Sathe, Olaf Schenk, and Ananth Grama. Fast parallel algorithms for graph similarity and matching. *Journal of Parallel and Distributed Computing*, 74(5):2400–2410, May 2014.
- [32] Danaï Koutra, Hanghang Tong, and David Lubensky. BIG-ALIGN: Fast bipartite graph alignment. In *2013 IEEE 13th International Conference on Data Mining*. IEEE, December 2013.
- [33] Mehmet Koyutürk, Yohan Kim, Umut Topkara, Shankar Subramaniam, Wojciech Szpankowski, and Ananth Grama. Pairwise alignment of protein interaction networks. *Journal of Computational Biology*, 13(2):182–199, March 2006.
- [34] O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes, and N. Przulj. Topological network alignment uncovers biological function and phylogeny. *Journal of The Royal Society Interface*, 7(50):1341–1354, March 2010.
- [35] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, March 1955.
- [36] D. Anil Kumar, A. S. C. S. Sastry, P. V. V. Kishore, and E. Kiran Kumar. Indian sign language recognition using graph matching on 3d motion captured signs. *Multimedia Tools and Applications*, June 2018.
- [37] E Kiran Kumar, PVV Kishore, D Anil Kumar, and M Teja Kiran Kumar. Early estimation model for 3d–discrete indian sign language recognition using graph matching. *Journal of King Saud University-Computer and Information Sciences*, 2018.
- [38] Lorenzo Livi and Antonello Rizzi. The graph matching problem. *Pattern Analysis and Applications*, 16(3):253–283, August 2012.
- [39] Si Wei Lu, Ying Ren, and Ching Y. Suen. Hierarchical attributed graph representation and recognition of handwritten chinese characters. *Pattern Recognition*, 24(7):617–632, January 1991.
- [40] Bruno T Messmer. *Efficient graph matching algorithms*. PhD thesis, 1995.
- [41] B.T. Messmer and H. Bunke. A new algorithm for error-tolerant subgraph isomorphism detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(5):493–504, May 1998.
- [42] R. Milo. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, October 2002.

- [43] Mark Newman. *Networks: An Introduction*. OUP Oxford, 2010.
- [44] Behnam Neyshabur, Ahmadreza Khadem, SoMaye Hashemifar, and Seyed Shahriar Arab. Netal: a new graph-based method for global alignment of protein–protein interaction networks. *Bioinformatics*, 29(13):1654–1662, 2013.
- [45] Soonyong Park, Sung-Kee Park, and Martial Hebert. Fast and scalable approximate spectral matching for higher order graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):479–492, March 2014.
- [46] R. Patro and C. Kingsford. Global network alignment using multiscale spectral signatures. *Bioinformatics*, 28(23):3105–3114, October 2012.
- [47] Hang TT Phan and Michael JE Sternberg. Pinalog: a novel approach to align protein interaction networks—implications for complex detection and function prediction. *Bioinformatics*, 28(9):1239–1245, 2012.
- [48] N. Przulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, January 2007.
- [49] Alex Radu and Michael Charleston. Node fingerprinting: an efficient heuristic for aligning biological networks. *Journal of Computational Biology*, 21(10):760–770, 2014.
- [50] Kaspar Riesen and Horst Bunke. Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision Computing*, 27(7):950–959, June 2009.
- [51] J. Rocha and T. Pavlidis. A shape analysis model with applications to a character recognition system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(4):393–404, April 1994.
- [52] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, September 2010.
- [53] Alberto Sanfeliu and King-Sun Fu. A distance measure between attributed relational graphs for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):353–362, May 1983.
- [54] Vikram Saraph and Tijana Milenković. Magna: Maximizing accuracy in global network alignment. *Bioinformatics*, 30(20):2931–2940, July 2014.
- [55] Francesc Serratosa. Fast computation of bipartite graph matching. *Pattern Recognition Letters*, 45:244–250, August 2014.
- [56] Linda G. Shapiro and Robert M. Haralick. Structural descriptions and inexact matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(5):504–519, September 1981.

- [57] R. Sharan, S. Suthram, R. M. Kelley, T. Kuhn, S. McCuine, P. Uetz, T. Sittler, R. M. Karp, and T. Ideker. Conserved patterns of protein interaction in multiple species. *Proceedings of the National Academy of Sciences*, 102(6):1974–1979, February 2005.
- [58] Roded Sharan and Trey Ideker. Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, 24(4):427–433, April 2006.
- [59] R. Singh, J. Xu, and B. Berger. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proceedings of the National Academy of Sciences*, 105(35):12763–12768, August 2008.
- [60] Rohit Singh, Jinbo Xu, and Bonnie Berger. Pairwise global alignment of protein interaction networks by matching neighborhood topology. In *Lecture Notes in Computer Science*, pages 16–31. Springer Berlin Heidelberg.
- [61] J. R. Ullmann. An algorithm for subgraph isomorphism. *Journal of the ACM*, 23(1):31–42, January 1976.
- [62] S. Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):695–703, 1988.
- [63] Huanran Wang, Hui He, and Weizhe Zhang. Demadroid: Object reference graph-based malware detection in android. *Security and Communication Networks*, 2018:1–16, May 2018.
- [64] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440, 1998.
- [65] Wikipedia contributors. Maximum common induced subgraph — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=Maximum-common-induced-subgraph&oldid=749430208>, 2016. [Online; accessed 30-August-2018].
- [66] Weiguo Zheng, Lei Zou, Xiang Lian, Dong Wang, and Dongyan Zhao. Efficient graph similarity search over large graph databases. *IEEE Transactions on Knowledge and Data Engineering*, 27(4):964–978, April 2015.

## INDEX

edge density, 8

global clustering coefficient, 8