Title: Biological graph dissimilarity characterization using graph theory

Date: May 07, 2014 11:55 AM

URL: http://pirsa.org/14050043

Abstract: Many biological data sets and relationships can be modeled as graphs. Understanding how structure of these graphs relates to biological function is essential for understanding underlining mechanisms of disease and for aiding drug discoveries. Vertices of biological graphs represent individual entities such as genes and proteins. Edges represent the relationship between two cellular components such as physical and functional interactions. A challenging problem in the post-genomic era is graph comparisons as they are large typed complex and evolving. Comparing graph structures helps to gain insights into the underlying signaling mechanisms and treatments for complex diseases. With technological advancement biological data will continue to grow and so will the size and complexity of graphs.Large graph comparisons are computationally intensive as they involve the subgraph isomorphism problem which is NP-complete. Therefore graph comparison algorithms need to be efficient scalable and be able to systematically capture biologically meaningful graph structure differences. Efficient graph comparison algorithms are necessary for many types of biological graphs e.g. protein-protein interaction drug-target microRNA-gene gene-regulatory and co-expression graphs. Furthermore graph comparison algorithms are extremely useful for many applications such as comparing graphs characterizing different diseases representing different cancer subtypes or different drug treatment responses. There are two main categories of graph properties used for comparing biological graphs global graph properties and local graph properties. Global graph properties study the overall graph while local graph properties focus on local structures of the graph. Our objective is to develop an efficient scalable graph comparison algorithm such that graph structure differences between any two states can be obtained systematically. We achieve the objective in two steps. First we propose an algorithm such that graph structure differences are systematically obtained and verified that the differences are biologically meaningful. Then we develop a heuristic to improve upon the proposed algorithm in the first step in terms of efficiency and scalability. While our approaches are generic we apply it on non-small cell lung cancer data sets. The non-small cell lung cancer datasets are used to construct normal and tumor co-expression graphs. Global graphs properties do not contain the detail needed to capture the structural characteristics of biological graphs thus we used a local property graphlets. Graphlets are all non-isomorphic connected induced graphs on a specific number of vertices. By definition graphlets have the ability to capture all the local structures on a certain number of vertices. Results showed that our graphlet approach returns graph structure differences between normal and tumor conditions that correspond to biological knowledge. We then introduce a heuristic to identify areas that are likely to be different between the normal and tumor graph and perform graph comparisons on the identified areas only. The heuristic was able to achieve interesting results that were successfully validated in vitro.

Biological graph dissimilarity characterization using graph theory

Compute Ontario Research Day
May 7, 2014

Serene Wong

YORK UNIVERSITE UNIVERSITY

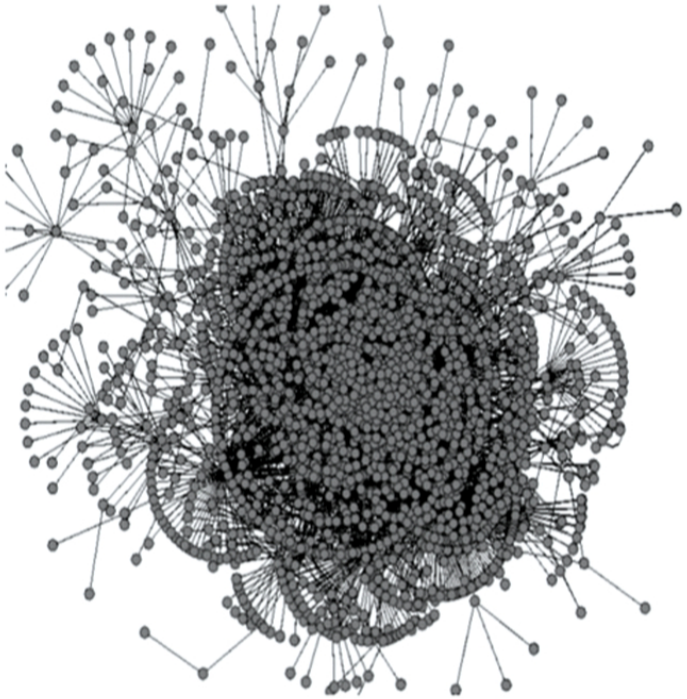
UofT

University Health Network

Biological graphs

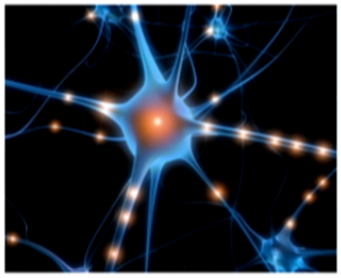
- Traditionally, individual cellular components and their functions are studied
- Most biological functions are due to interactions between different cellular constituents
- Various graphs have emerged

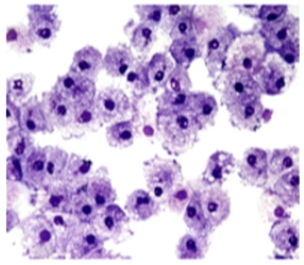
E.g. of vertices: genes, proteins
E.g. of edges: physical, functional interactions



Gene expression studies

In general, each cell in the body has the same DNA

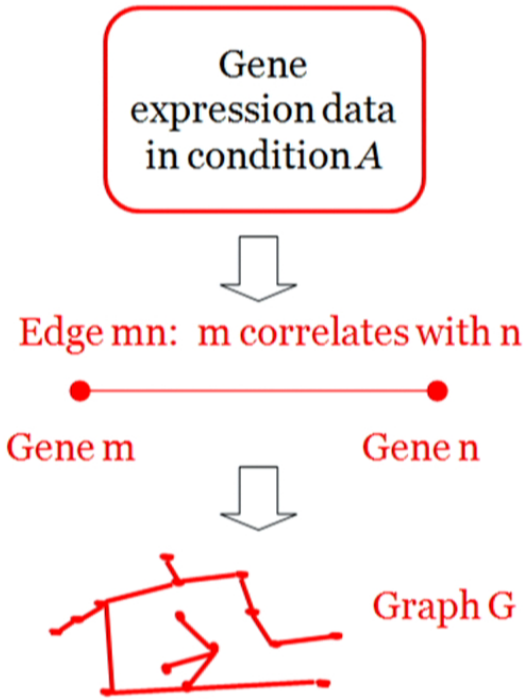


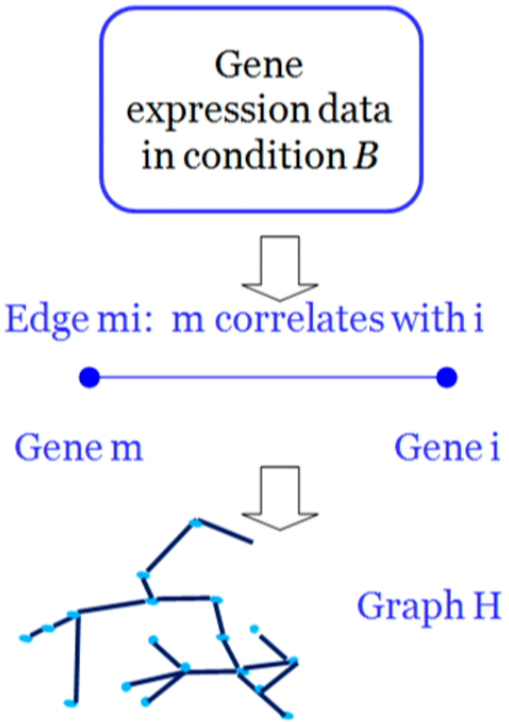


Different type of cells - difference is in the subset of genes that a cell expresses; proteins and their interactions

- Different responses to stimuli can also lead to expressing different subsets of genes
- Gene expression studies enable the understanding of the mechanism in the molecular level

Representing as graphs



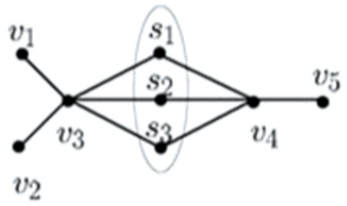


Example of graph structure & biological function

Graph structure

- If 2 vertices have the same neighborhood, then they are *siblings*

$$N(s_1) = N(s_2) = N(s_3) = \{v_3, v_4\}$$



Siblings: $s_1, s_2,$ and s_3.

Partial Fig. 1B of Functional topology in a network of protein interactions. (*Pržulj et al., 2004*)

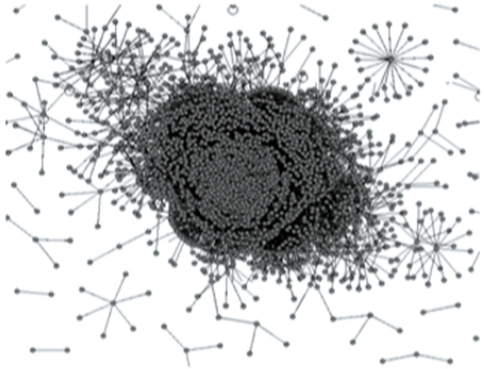
Biological function

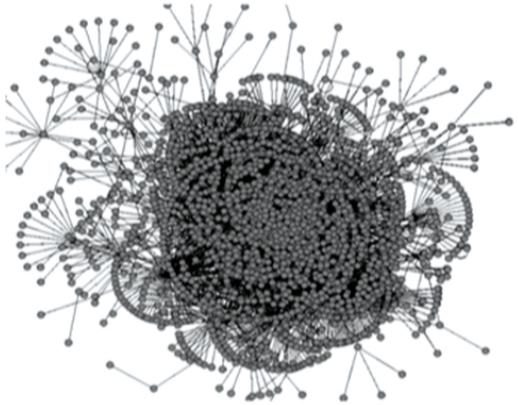
- A protein is viable if its mutation does not cause lethality of the cell
- Viable proteins were more frequent in the group of vertices that belonged to the sibling group

(*Pržulj et al., 2004*)

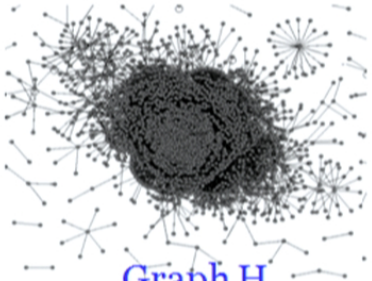
Graph comparisons

Comparing graph structures helps to gain insights into the underlying mechanisms and treatments for complex diseases

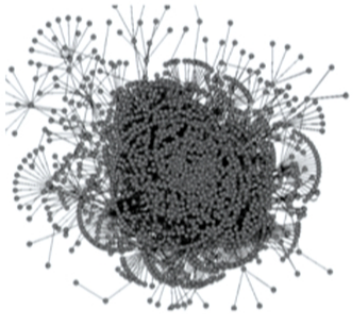




Objective



Graph H

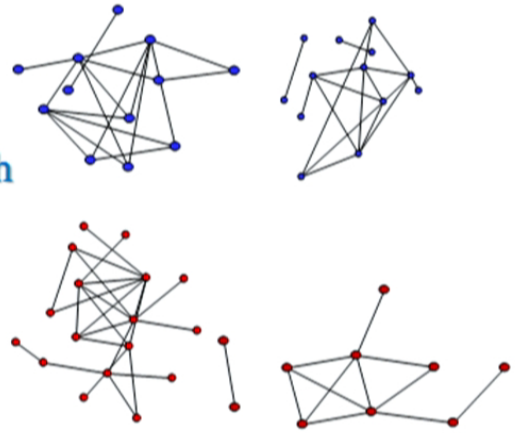


Graph G

Graph comparison algorithm

Efficient and scalable

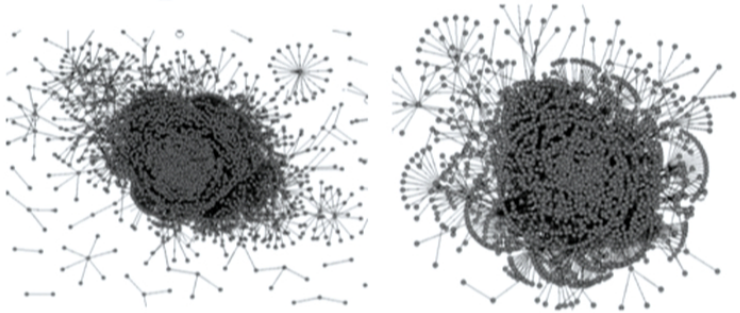
Systematically obtain graph structure differences



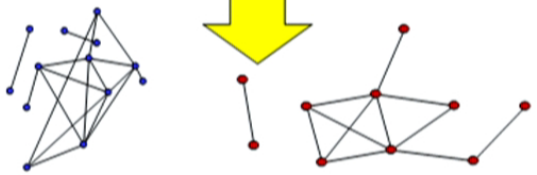
Graph structure differences

Method

Step 1: Exhaustive Search

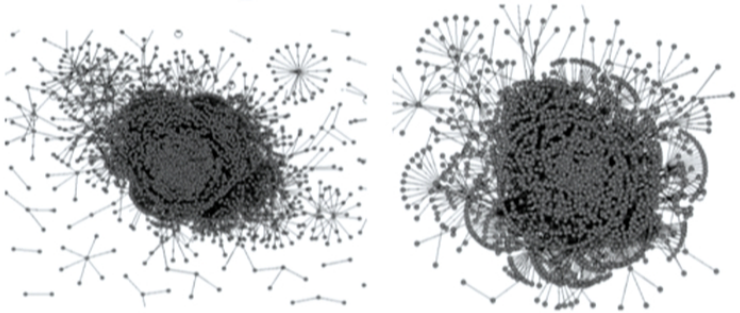


Graph G Graph H

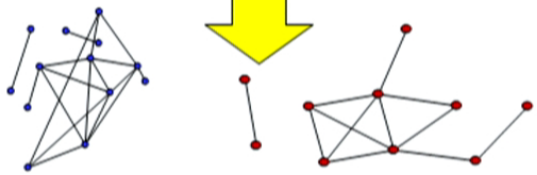


1. Graph structure differences are systematically obtained
2. Verified that the differences are biologically meaningful

Step 2: Heuristic



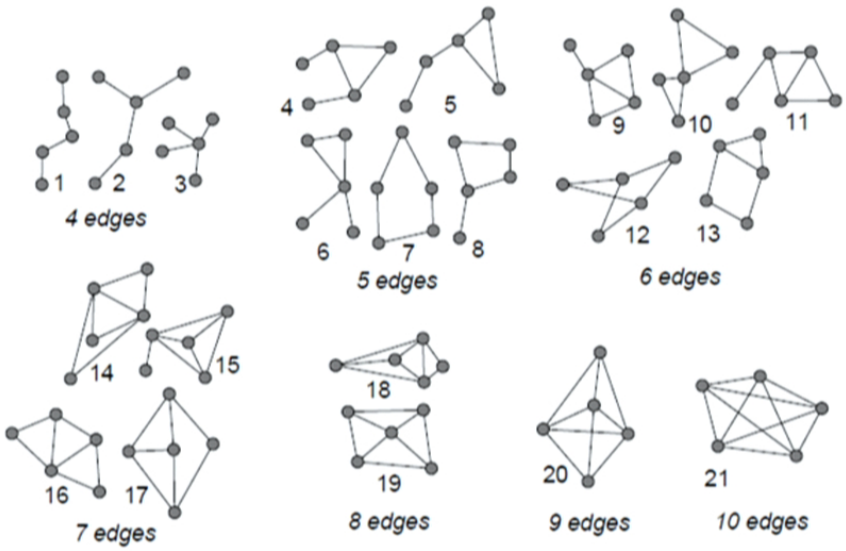
Graph G Graph H



Improve the efficiency

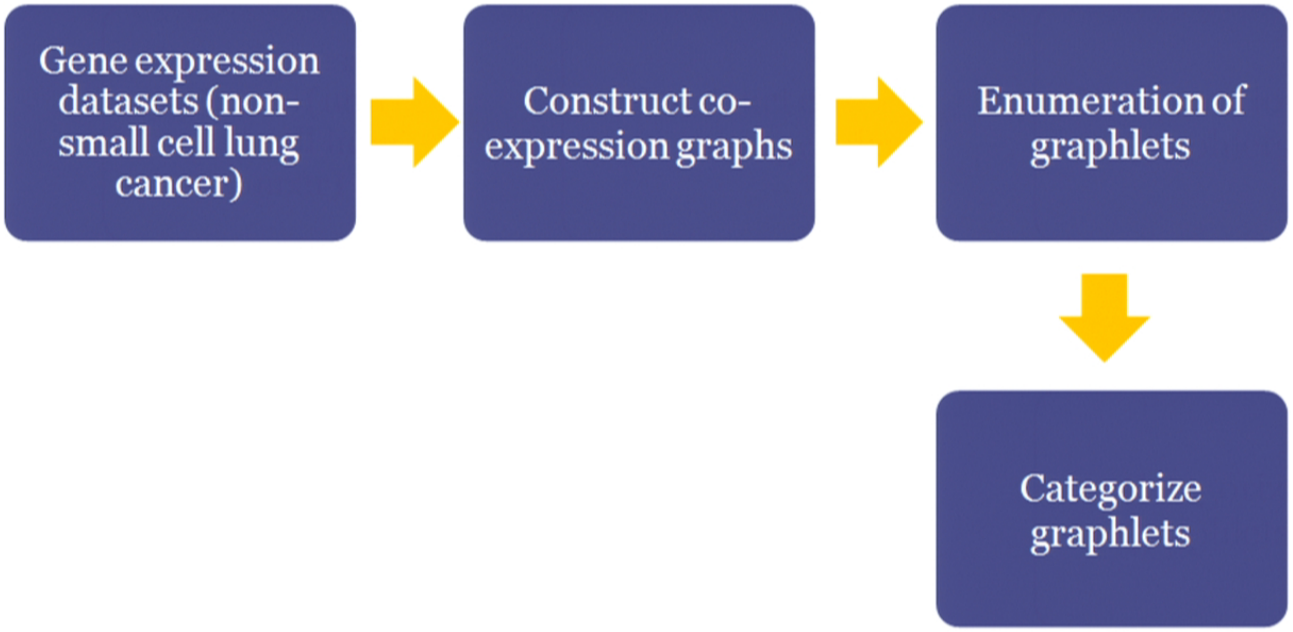
Graphlets

Graphlets: all non-isomorphic, connected induced graphs on a certain number of vertices *(Pržulj et al., 2004)*

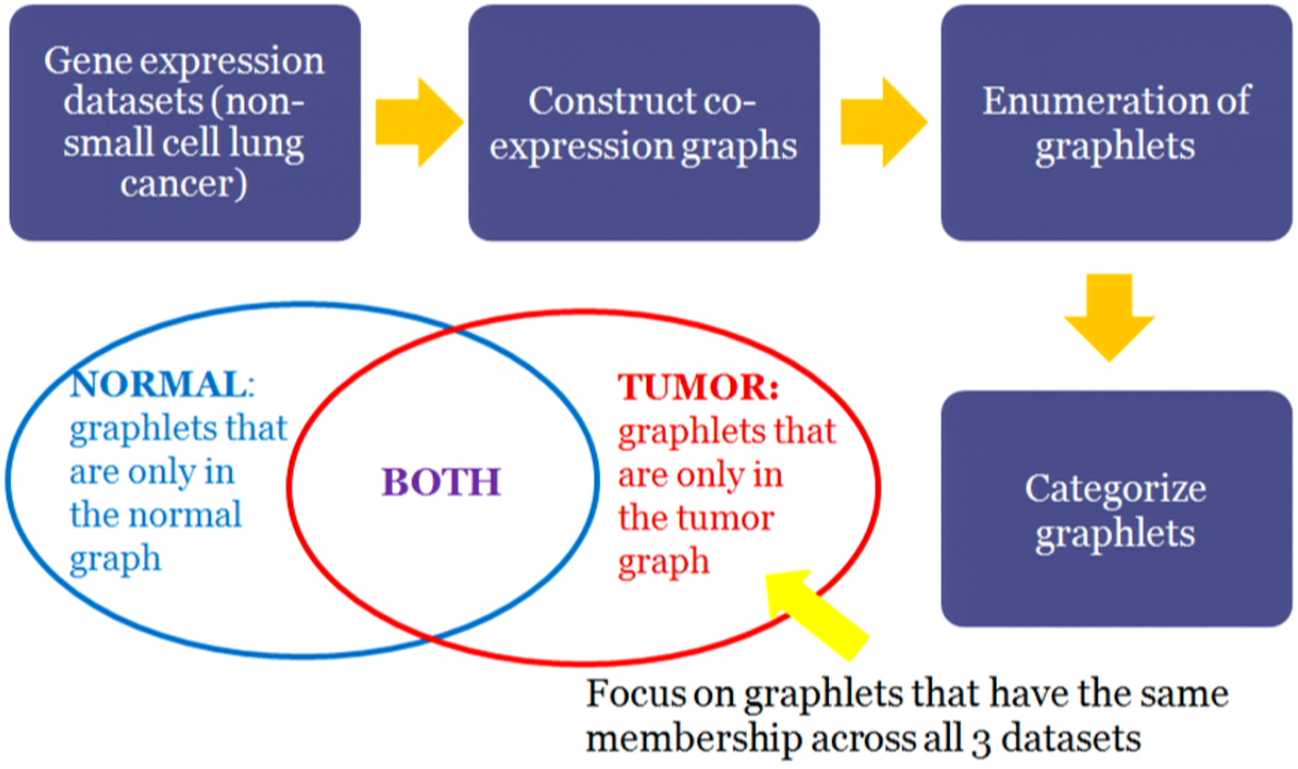


All twenty-one 5-node graphlets. All non-isomorphic, connected, induced graphs on 5 vertices.

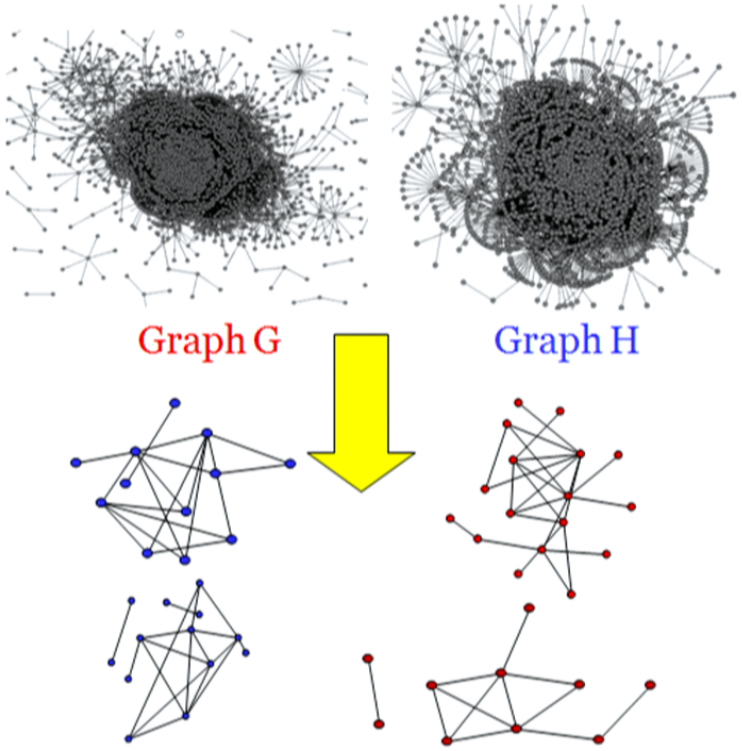
Step 1: Graphlet approach



Step 1: Graphlet approach



Results for the graphlet approach

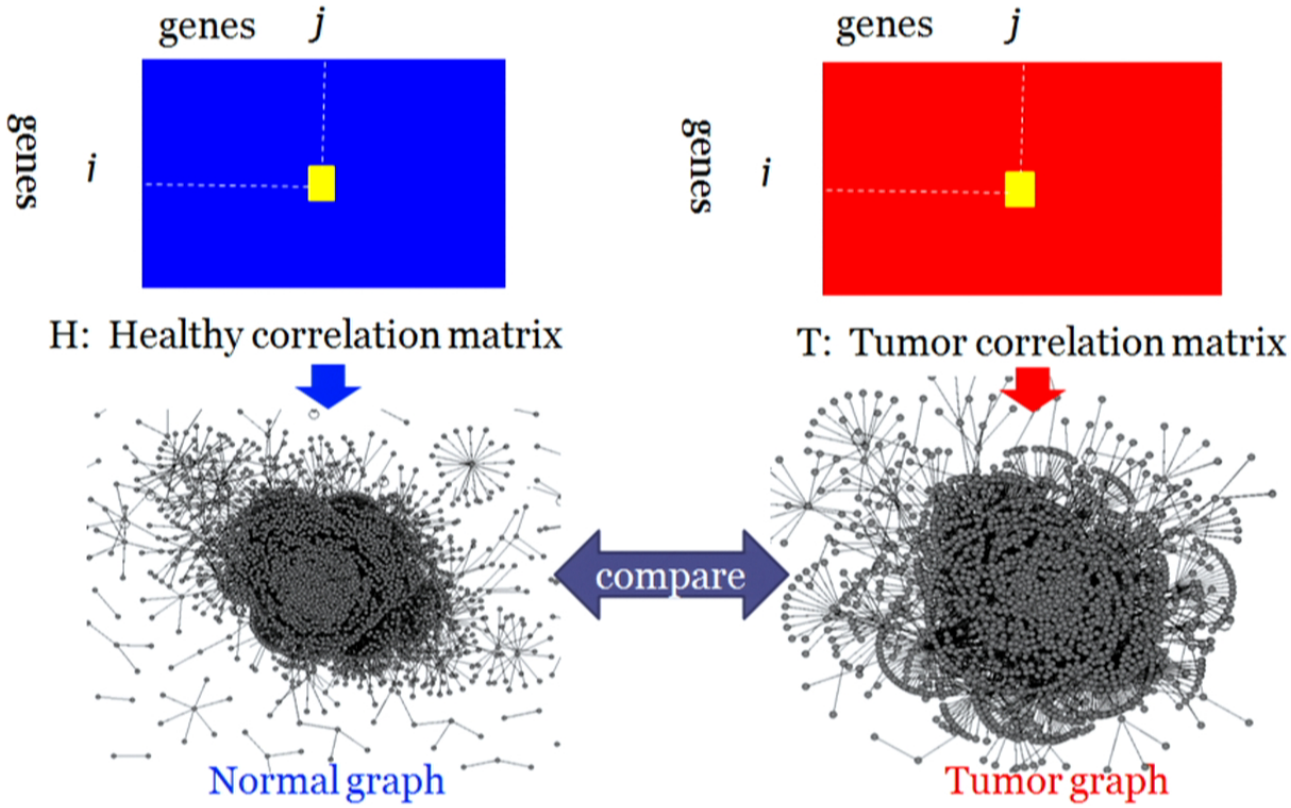


Graph structure differences

Biologically meaningful

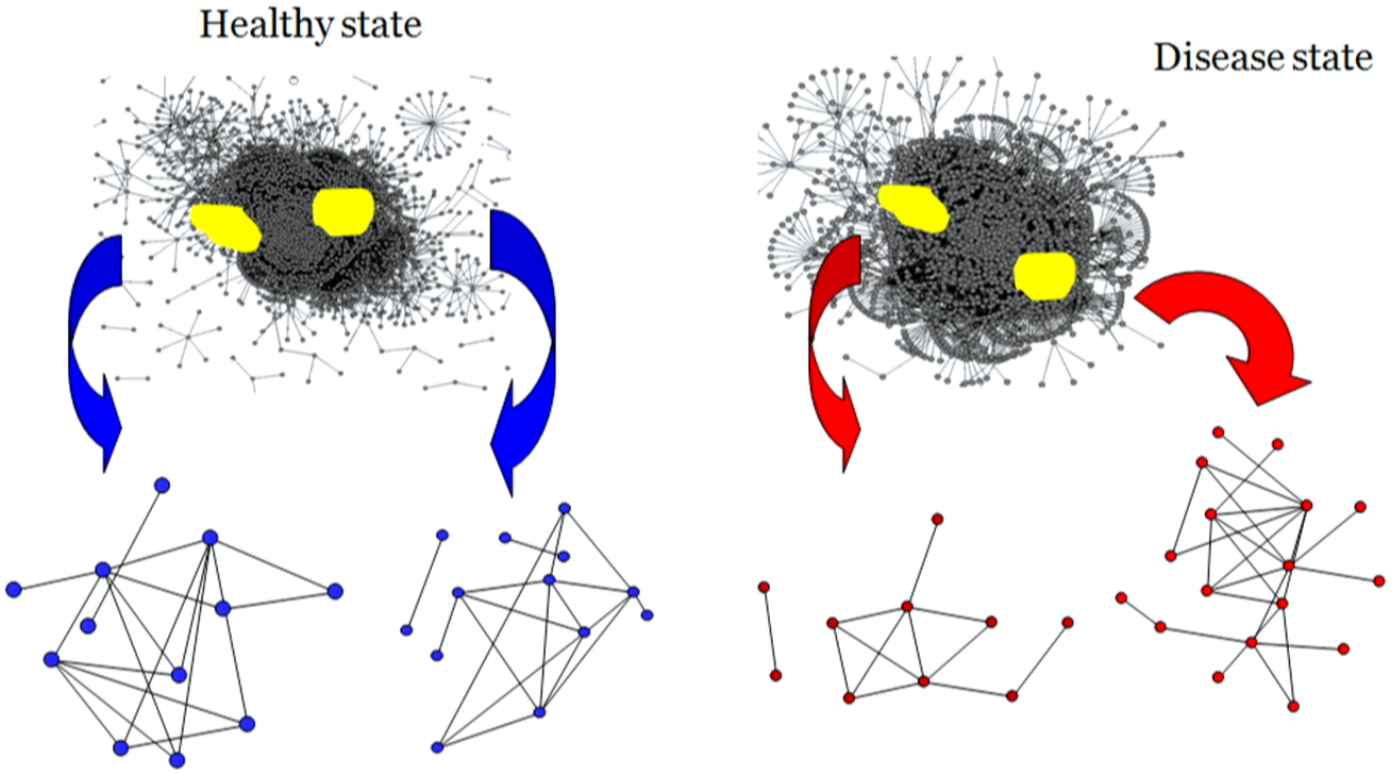
- biological process - "regulation of lymphocyte activation"
  - evading immune destruction is an emerging hallmark of cancer
- enriched in protein-protein interactions
- contains genes that are promising therapeutic targets in other cancers

Step 1: Graphlet approach

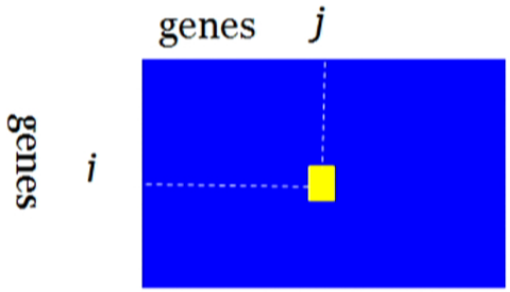


Step 2

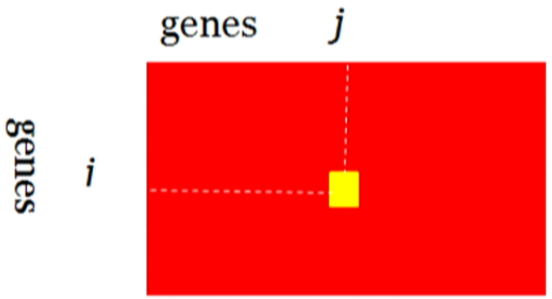
Identify important areas for comparing graphs, so that graphlet enumeration is performed only on important areas



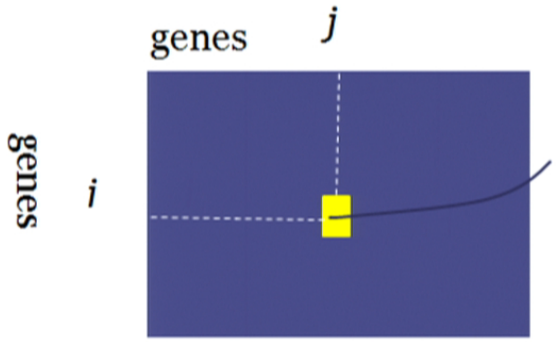
Step 2: The differential correlation graph (DCG) approach



H: Healthy correlation matrix

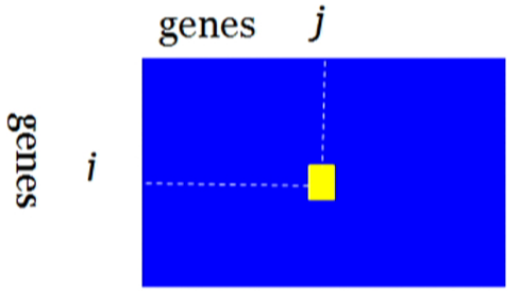


T: Tumor correlation matrix

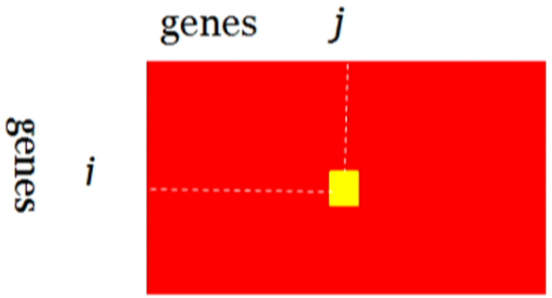


D: Absolute correlation difference matrix

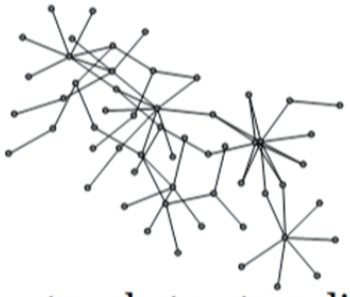
The differential correlation graph (DCG) approach



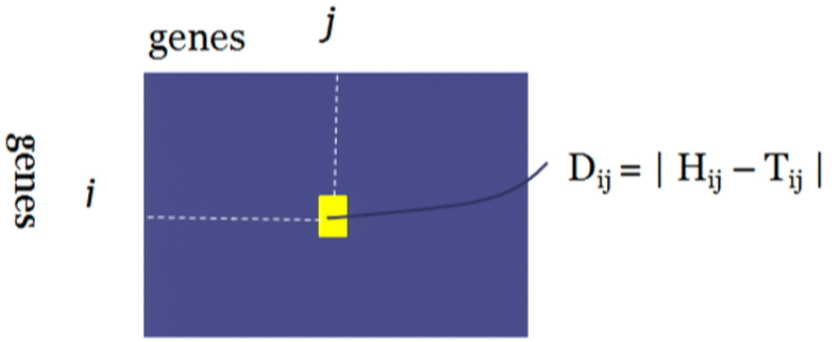
H: Healthy correlation matrix



T: Tumor correlation matrix



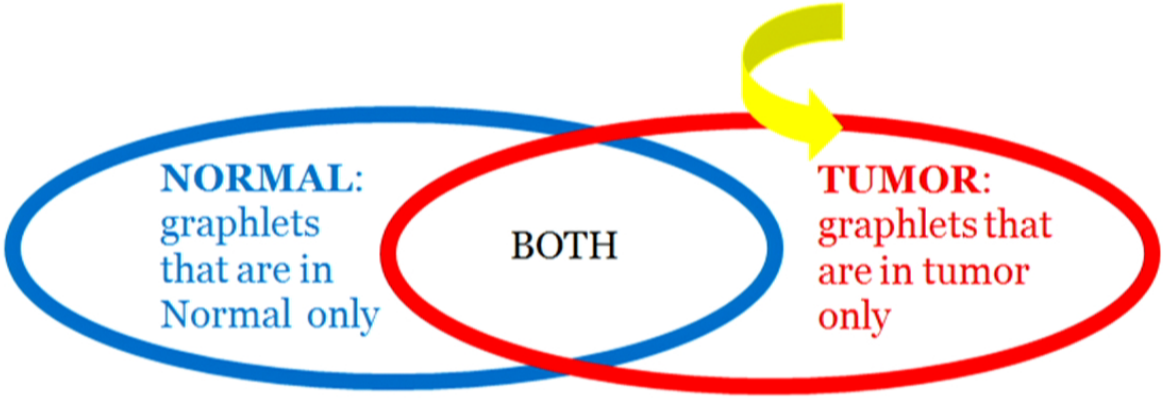
Obtain network structure differences by using neighborhoods of DCGs



D: Absolute correlation difference matrix

Benchmark - the all3 category

all3: graphlets that have the same membership across all 3 datasets



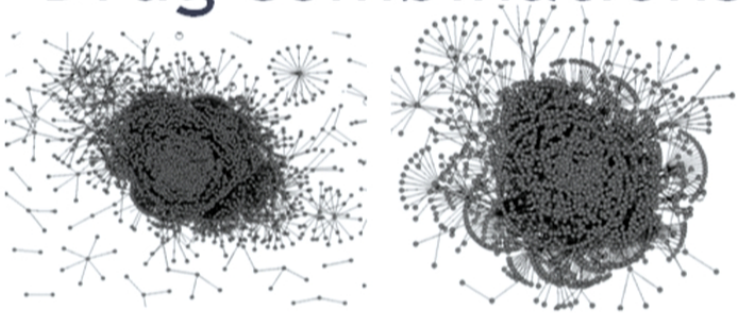
Results for the DCG approach

DCG	$\|V(DCG)\|$	% Node wrt N	% Node wrt T	$Approxall3_{DCG}$	$Accall3_{DCG}$
hou001	179	38.66	43.03	323	100
hou002	265	57.24	63.70	323	100
hou003	328	70.84	78.85	323	100
landi001	183	57.01	49.46	320	99.07
landi002	276	85.98	74.59	322	99.69
landi003	350	109.03	94.59	322	99.69
su001	186	43.97	42.96	310	95.98
su002	300	70.92	69.28	323	100.00
su003	379	89.60	87.53	323	100.00

Results for the all3 category for the *DCG* approach

All3 category is very important because all 3 datasets picked up these graphlets as graphlets that differed between the normal and tumor condition

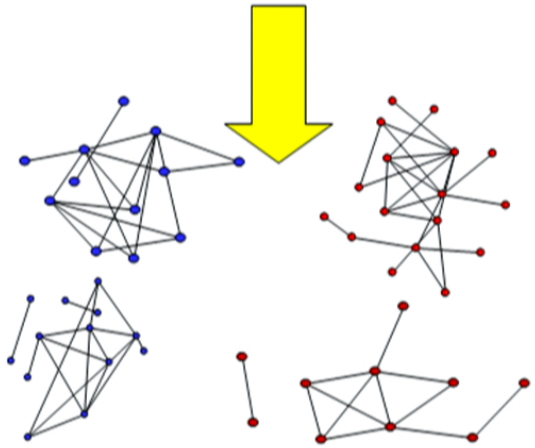
Drug combinations



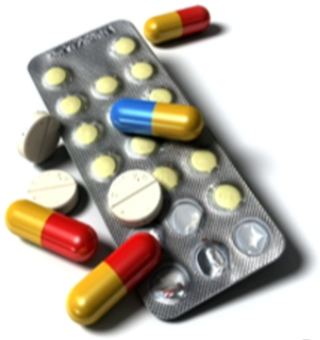
Healthy state

Disease state

Biological experiments



Graph structure differences



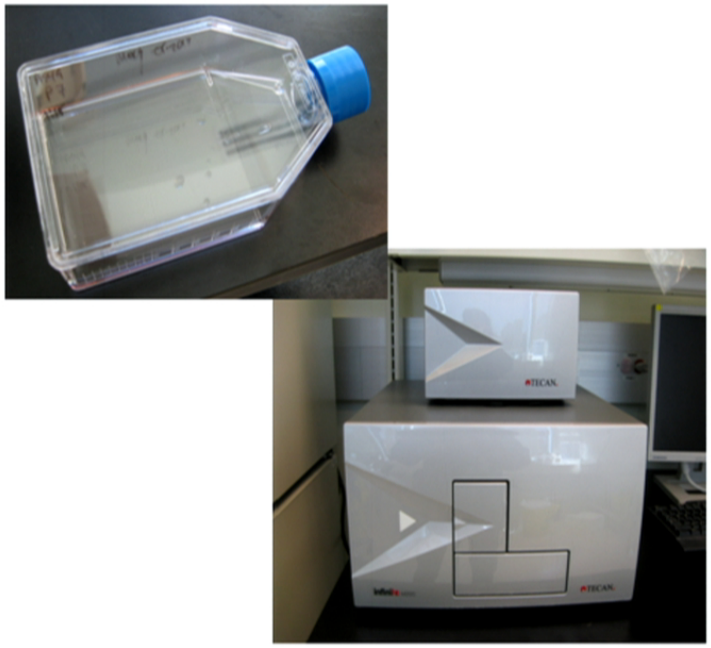
Drug combinations

Biological validation

Compare drug combination with the individual drugs

For all 3 cell lines, for all 3 drug combinations

1. cell viability is lowest for the predicted drug combinations
2. predicted drug combinations have lower cell viability than FDA approved drugs for non-small cell lung cancer



Conclusion

- Algorithms are generic
  - Non-small cell lung cancer datasets
- Graphlet approach
  - graph structure differences between normal and tumor conditions
  - correspond to biological knowledge
- DCG approach
  - achieve accurate results
  - successfully validated *in vitro*

Collaborators & Funding

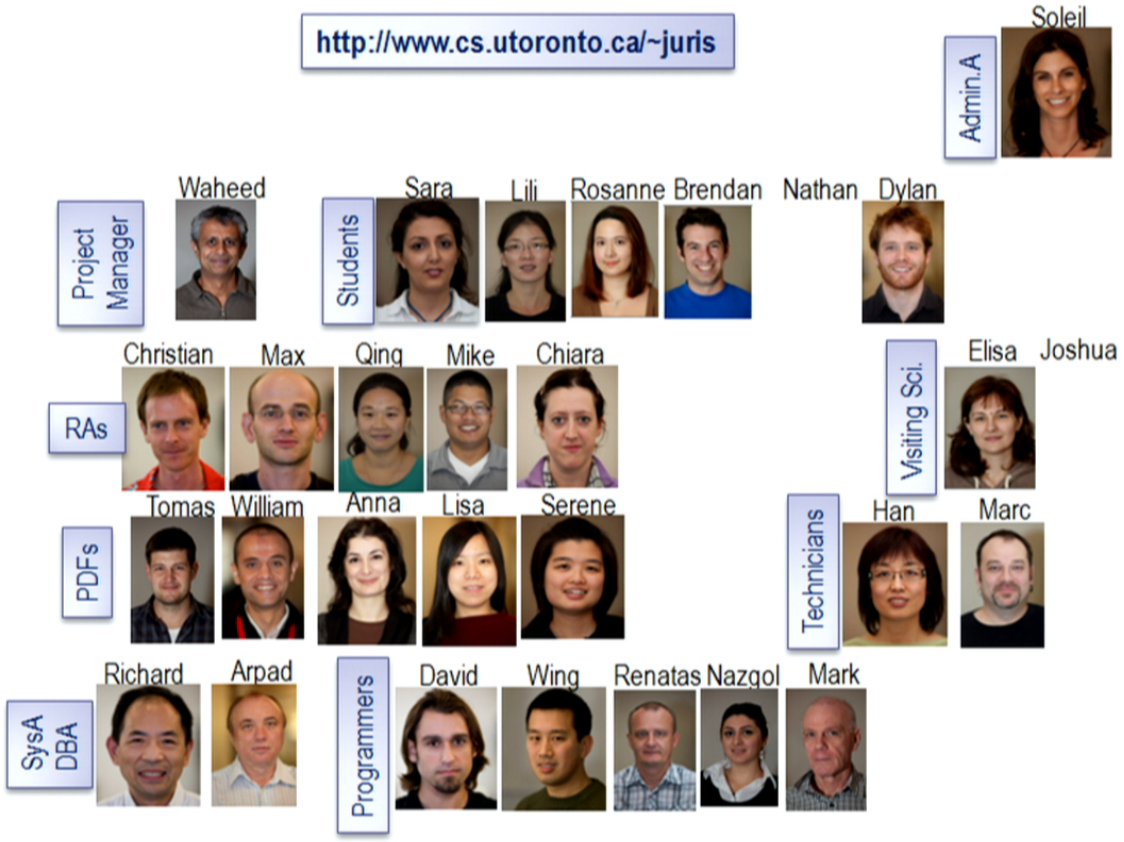
Lung cancer: Drs. Shepherd, Sound-Tsao, Lam, Reis, et al.
Head&neck cancer: Dr. Kamel-Reid
Pancreas cancer: Drs. Hedley, Reis
Ovarian Cancer: Drs. Oza, Mes-Masson, Jurisicova, Kaur, Kislinger, Clarke, et al.
Leukemia: Drs. Dick, Minden, Wong
Prostate cancer: Drs. Bristow, Fleshner

Drs. Stagljar, Maestro, Mills, DeTitta, Luft, Snell, …
N. Cercone

University Health Network

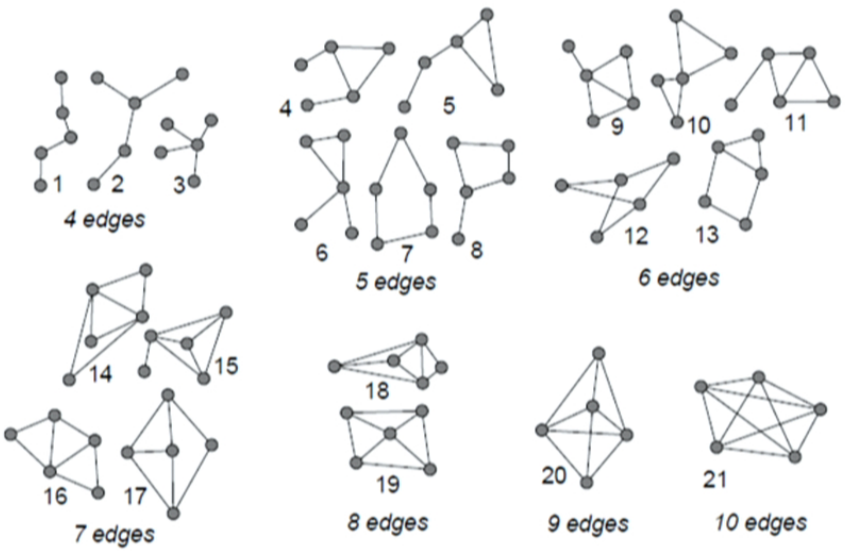
Ontario
Canada Research Chairs
Canada Foundation for Innovation
Fondation canadienne pour l'innovation
NSERC CRSNG
CIHR IRSC
THE FRIULI CANCER RESEARCH ACADEMIC EXCHANGE PROGRAM
Quorum
SIGMA-ALDRICH
lenovo
NVIDIA
world community grid.
Powered by IBM technology

http://www.cs.utoronto.ca/~juris



Graphlets

Graphlets: all non-isomorphic, connected induced graphs on a certain number of vertices *(Pržulj et al., 2004)*



All twenty-one 5-node graphlets. All non-isomorphic, connected, induced graphs on 5 vertices.