Data-Driven Medicine

Nataša Pržulj, PhD, MAE

ICREA Research Professor
Barcelona Supercomputing Center

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Overview

Medicine: complex world of inter-connected entities

1. **Motivation**
2. **New Methods – Examples:** mine inter-connected data
   i. **Single type of omics data: Molecular networks** → function, disease
   ii. **Multiple layers of heterogeneous data:**
      - **Patient-centered data integration → Precision medicine**
      - **Disease re-classification**
      - **Gene Ontology reconstruction**
      - **Network alignment**
3. **Vision**

Overview

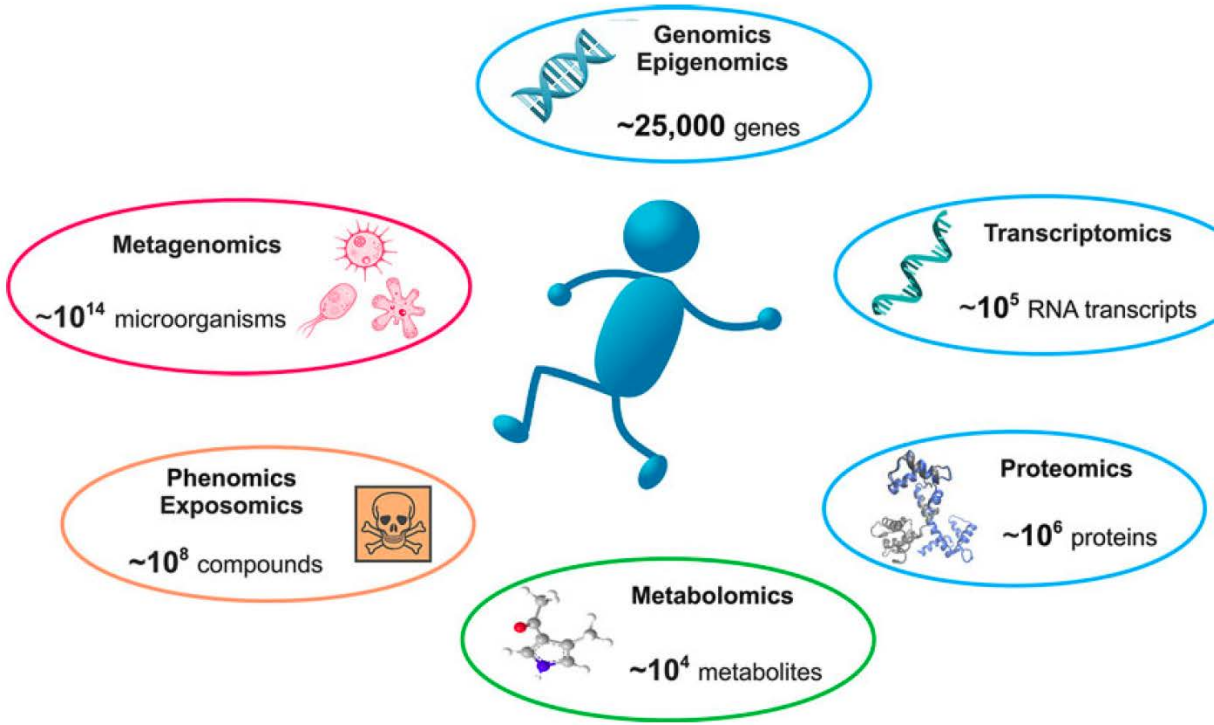
Medicine: complex world of inter-connected entities

1. **Motivation**
2. **New Methods – Examples:** mine inter-connected data
   i. **Single type of omics data: Molecular networks** → function, disease
   ii. **Multiple layers of heterogeneous data:**
      - **Patient-centered data integration → Precision medicine**
      - **Disease re-classification**
      - **Gene Ontology reconstruction**
      - **Network alignment**
3. **Vision**

1. Motivation
Medicine: complex world of inter-connected entities

Technological advances →
astounding harvest of various molecular and clinical data

Proteomics 2016, *16*, 741–758



REVIEW

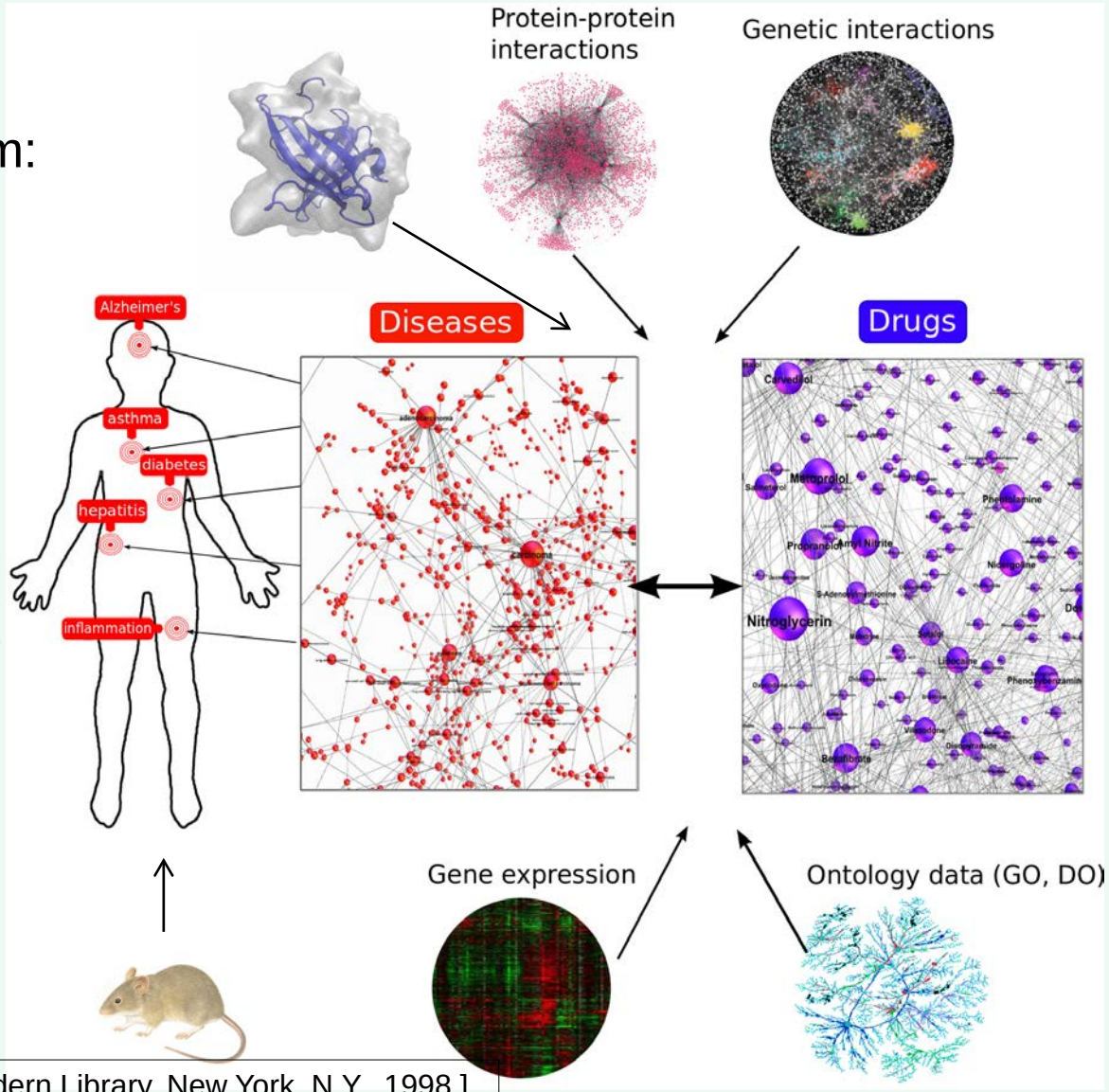
Integrative methods for analyzing big data in precision medicine

Vladimir Gligorijević, Noël Malod-Dognin and Nataša Pržulj

1. Motivation
Medicine: complex world of inter-connected entities

Data growth:

- Guided by empirical reductionism:
  - Striving to dissect a biological entity into its constituent parts
  - To better understand it
- However, knowing parts is not enough:
  - 1859 — Darwin[1] saw biology as a “tangled bank” with all its aspects interconnected
  - 1855 — Virchow[2]: all diseases involve changes in normal cells

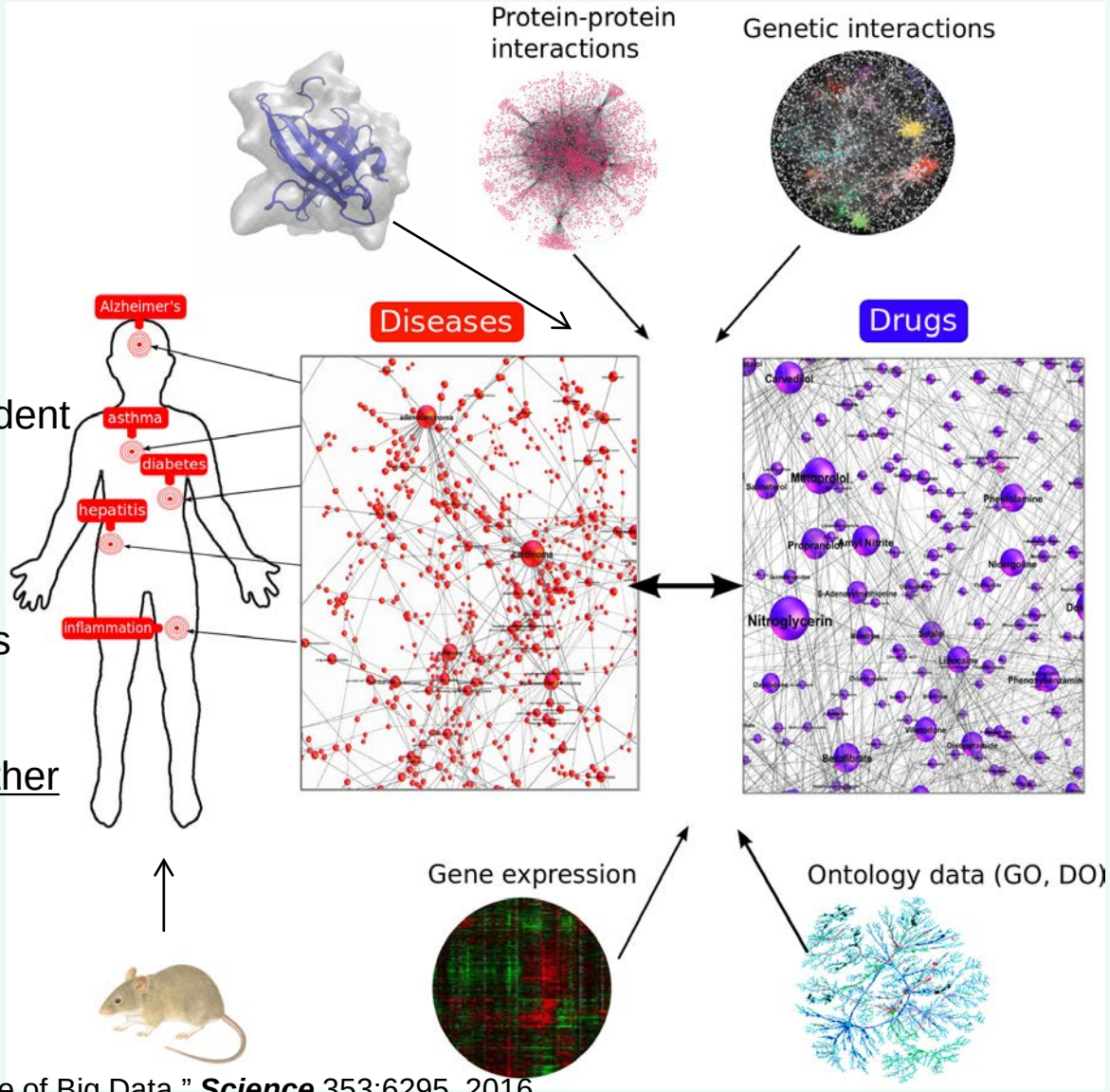


1. Darwin, C., On the origin of species, 1859 [Reprint, Modern Library, New York, N.Y., 1998.]
2. Virchow R., *Arch. Pathol. Anat. u. Physiol. u. klin. Med*. 8:3, 1855

1. Motivation
Medicine: complex world of inter-connected entities

Data growth about a cell:

- **Hit the wall of bio-complexity**
- Cells:
  - are not just loosely coupled arrangements of quasi-independent molecules
  - highly intricately and precisely integrated **networks** of **entities** and **interactions** within the cells and with the environment
  - Data types complement each other
  - Seek joint modeling and mining

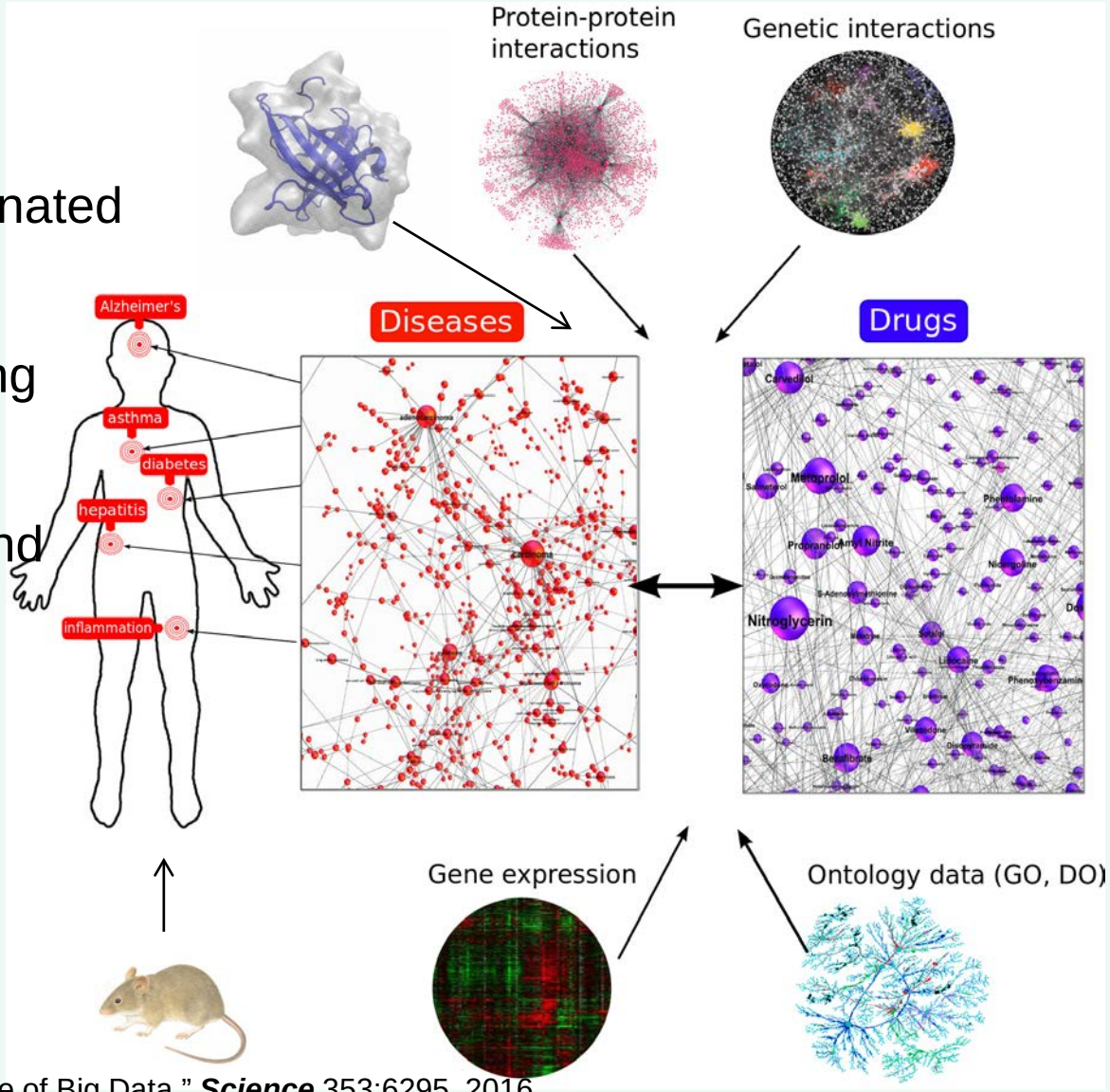


N. Przulj and N. Malod-Dognin, "Network analytics in the age of Big Data," ***Science*** 353:6295, 2016
C. R. Woese, A New Biology for a New Century, *Microbiology and Molecular Biology Reviews* 68(2):173-186, 2004

1. Motivation
Medicine: complex world of inter-connected entities

Time to:

- Replace the mostly reductionist molecular perspective that dominated the 20th century
- New and holistic view of the living world
- Required to explain biological and medical phenomena
- Biology's innate complexity



N. Przulj and N. Malod-Dognin, "Network analytics in the age of Big Data," ***Science*** 353:6295, 2016
C. R. Woese, A New Biology for a New Century, *Microbiology and Molecular Biology Reviews* 68(2):173-186, 2004

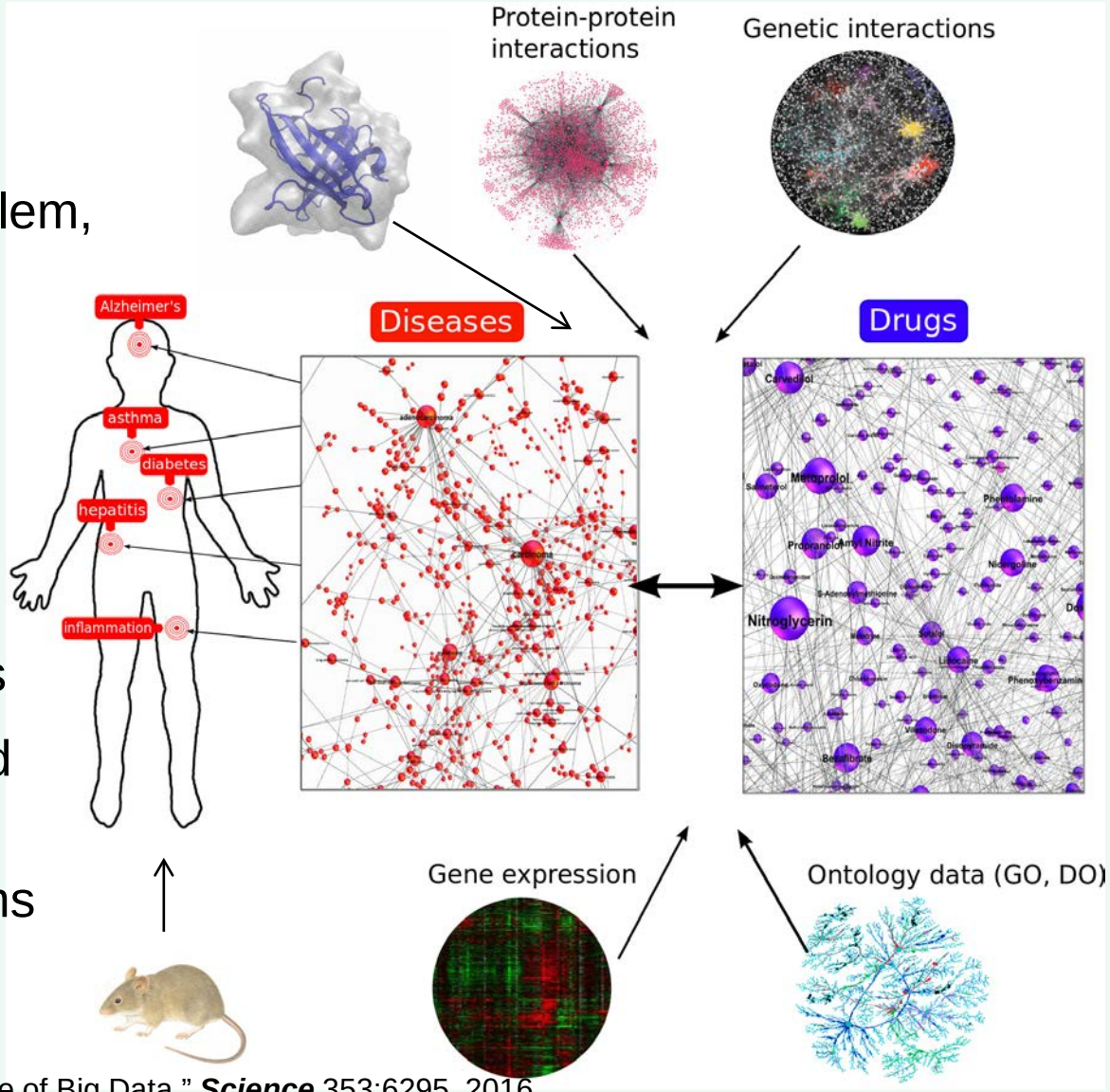
1. Motivation
Medicine: complex world of inter-connected entities

Requires:

- Establishing a perspective and framework not only for one problem, but for biology and medicine in general

A foremost challenge:

- **How to re-synthesize biology**
- Put the elements back into their complex, dynamic environments
- Connect them all within a unified framework
- Reformulate biological paradigms within the non-linear world



N. Przulj and N. Malod-Dognin, "Network analytics in the age of Big Data," ***Science*** 353:6295, 2016
C. R. Woese, A New Biology for a New Century, *Microbiology and Molecular Biology Reviews* 68(2):173-186, 2004

1. Motivation
Medicine: complex world of inter-connected entities

Vision:

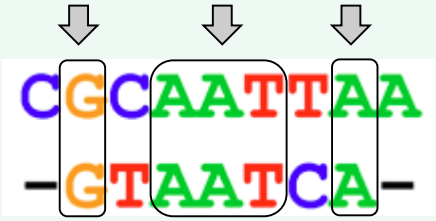
- Bridge this gap by developing a mathematically principled framework for integration of networked data
- Marry biomedical problems and data with algorithms from:
  - ML, such as NMTF
  - Mathematical non-linear optimization
  - Network science
  - Algebraic topology…
  - High-performance computing
- Propose modelling & computational advances that will link the medicine's:
  - reductionist past with its holistic future
- Enable
  - displacement of the dominant molecular representation of biology
  - by a new, integrative paradigm that is deeper, more comprehensive and inspiring

€2M **ERC Consolidator Grant** for 2018-2023
Title: "Integrated Connectedness for a New Representation of Biology"

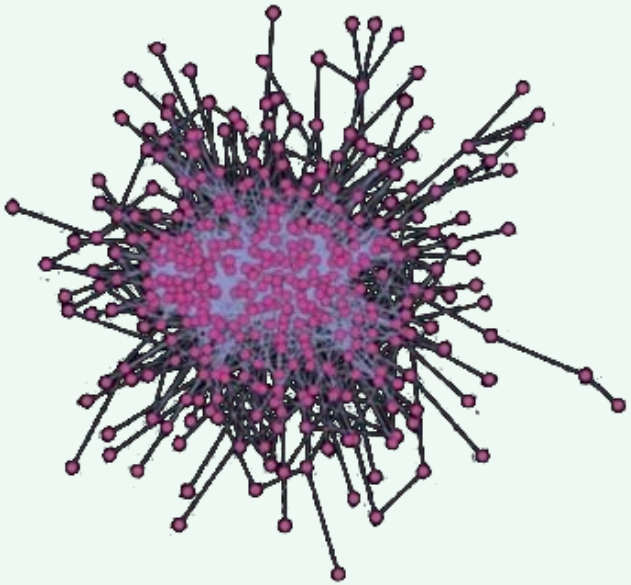
1. Motivation
Medicine: complex world of inter-connected entities

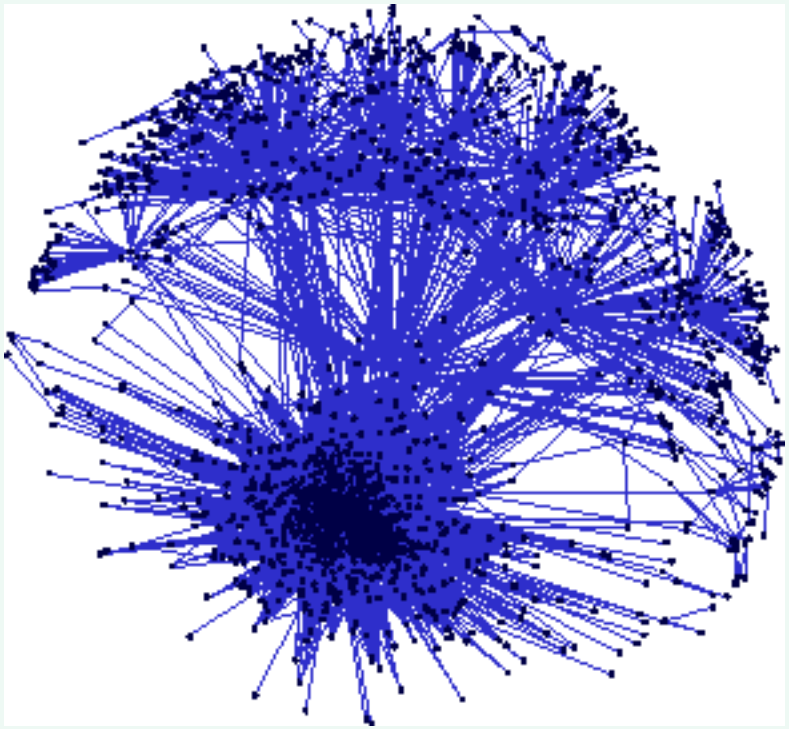
Computational challenges

- Need new tools to mine complex data systems
- Why?
  - Analysing sequences: “computationally easy” → still lacking
  - Analysing interconnected heterogeneous data: “computationally hard”



- Sophisticated methods carefully tuned to extract new knowledge from particular data

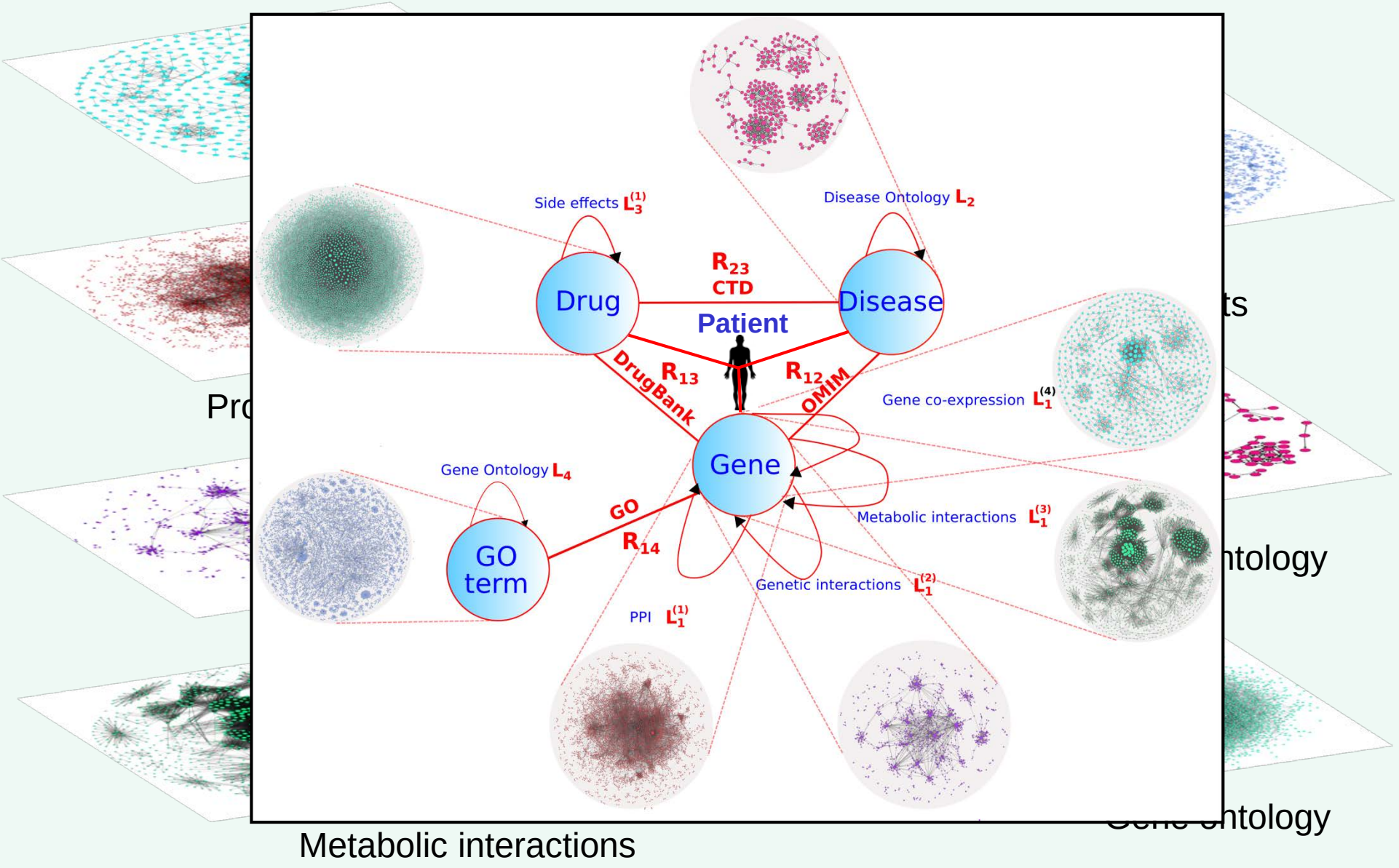




1. Motivation

Medicine: complex world of inter-connected entities

Computational challenges



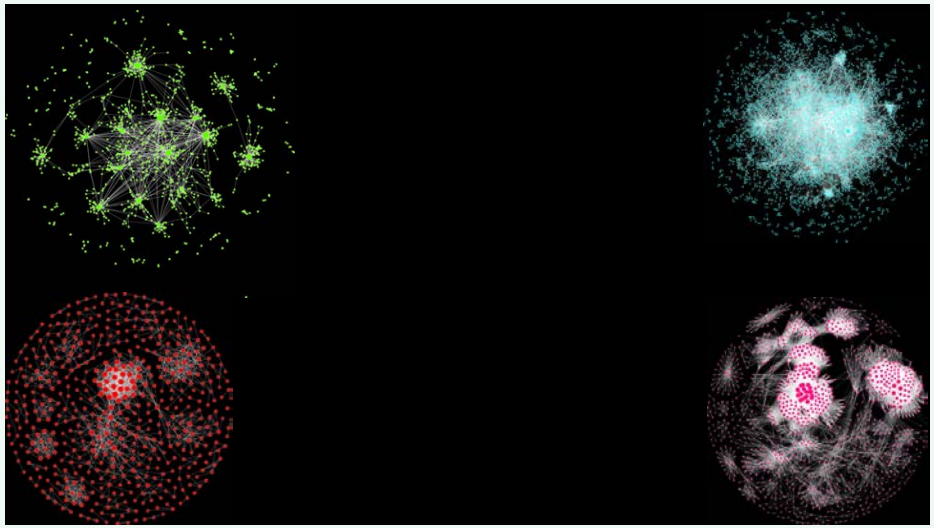
Overview

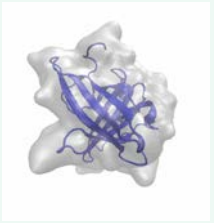
Medicine: complex world of inter-connected entities

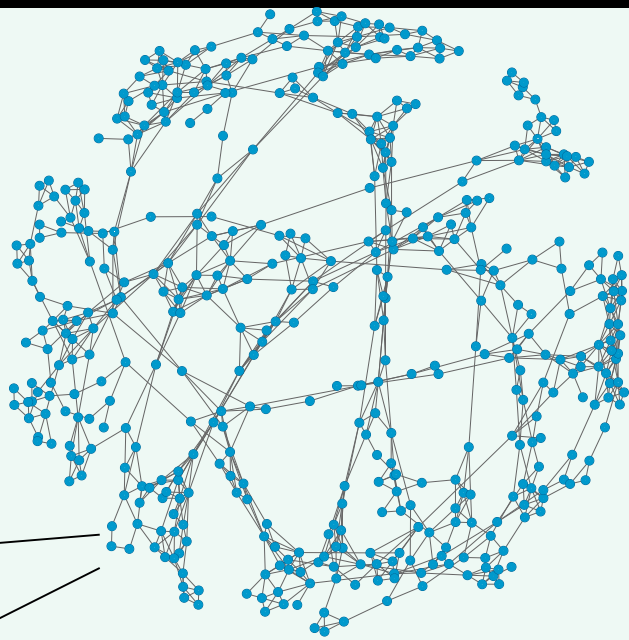
1. **Motivation**
2. **New Methods – Examples:** mine inter-connected data
   i. **Single layer of omics data: Molecular networks** → function, disease
   ii. **Multiple layers of heterogeneous data:**
      - **Patient-centered data integration → Precision medicine**
      - **Disease re-classification**
      - **Gene Ontology reconstruction**
      - **Network alignment**
3. **Vision**

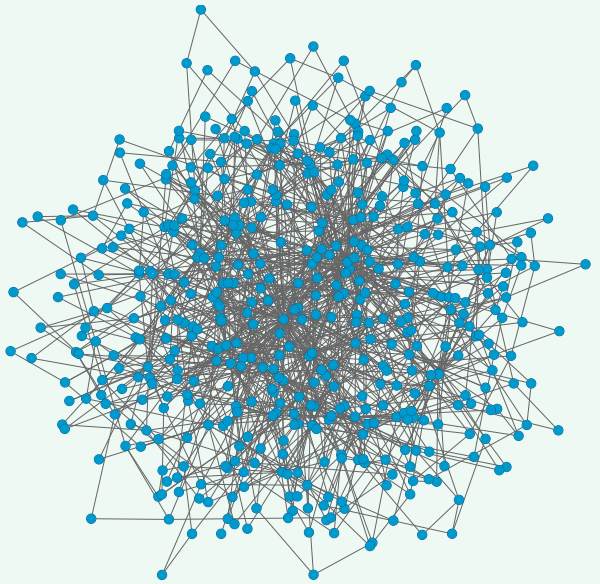
2. Novel Methods
Mine the Medical World of Inter-Connected Entities

i. Molecular Networks





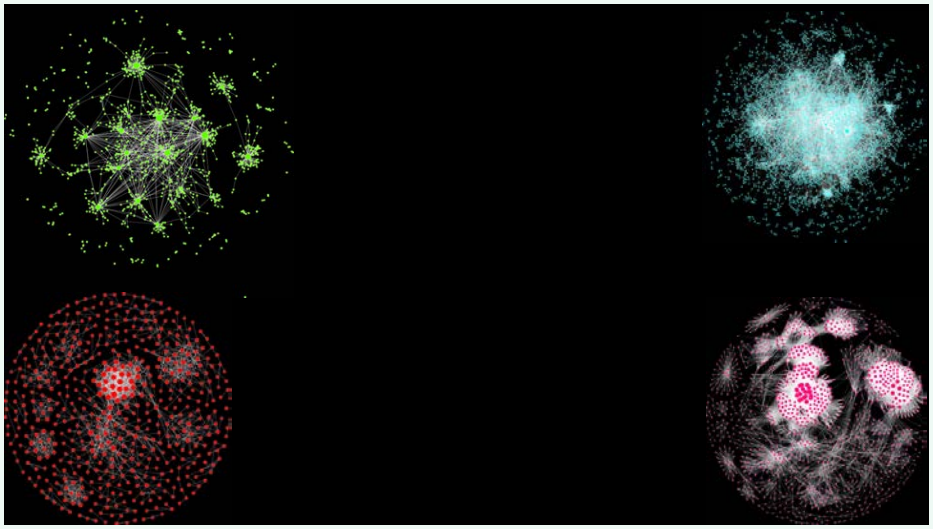


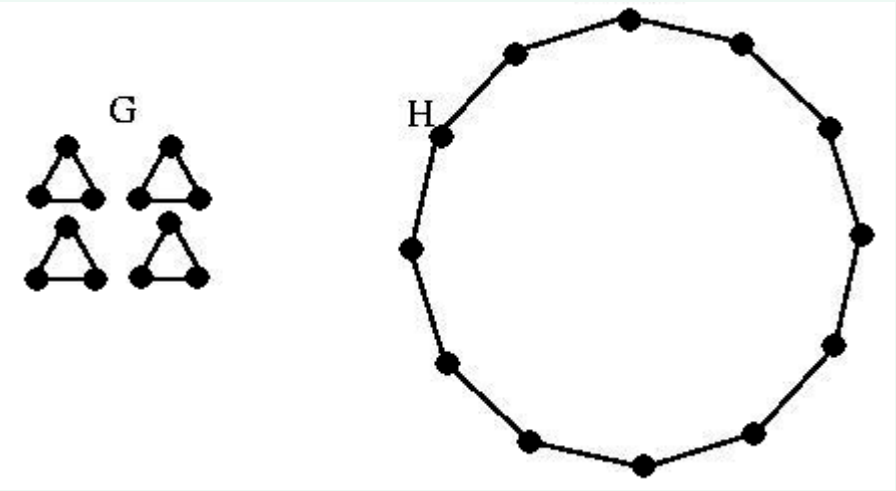


- The number of nodes
- The number of links
- Links of each node: *degree*
- Distribution of links (degrees)

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

i. Molecular Networks

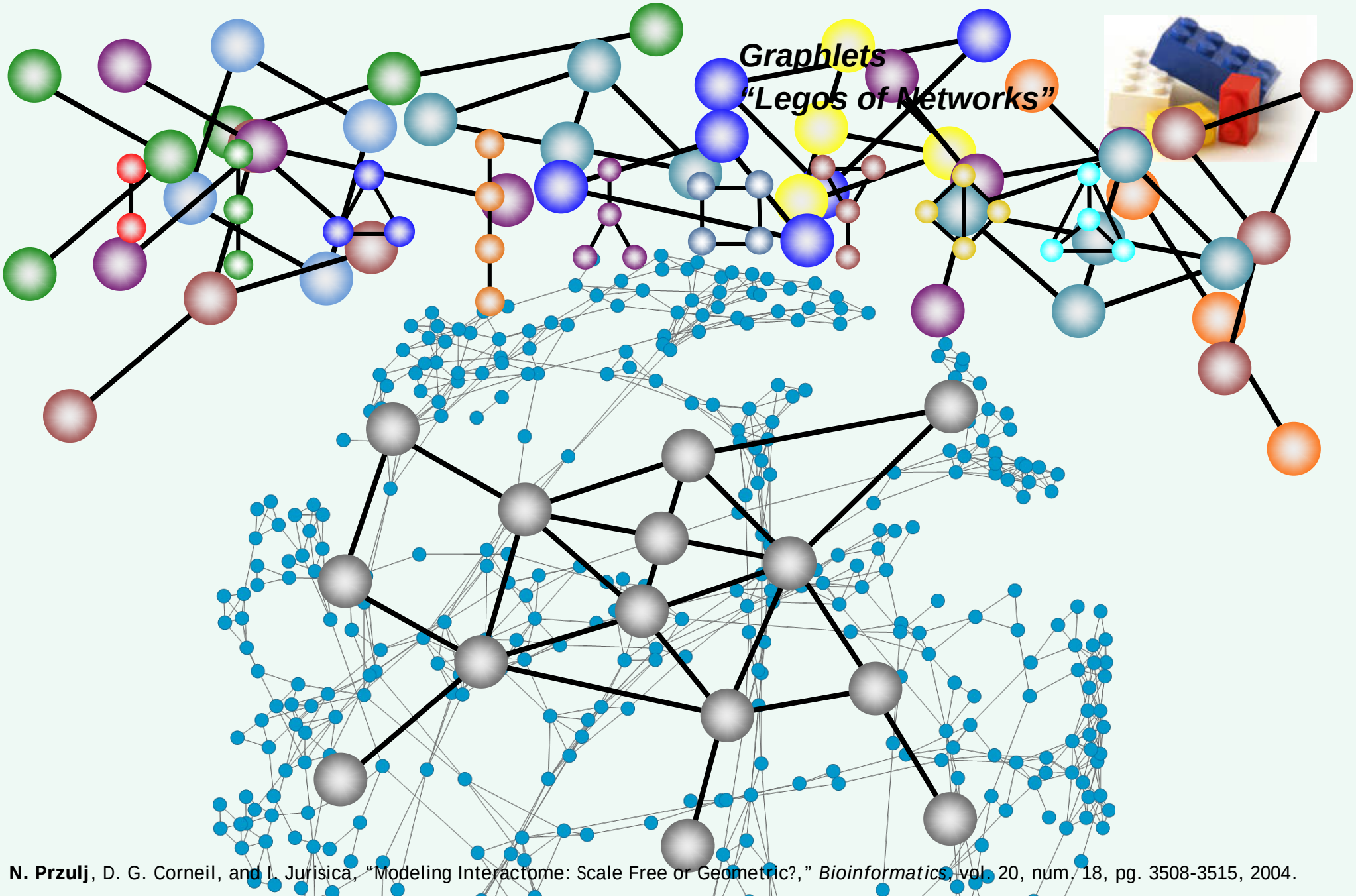




- The number of nodes
- The number of links
- Links of each node: *degree*
- Distribution of links (degrees)

2. Novel Methods

Mine the Medical World of Inter-Connected Entities



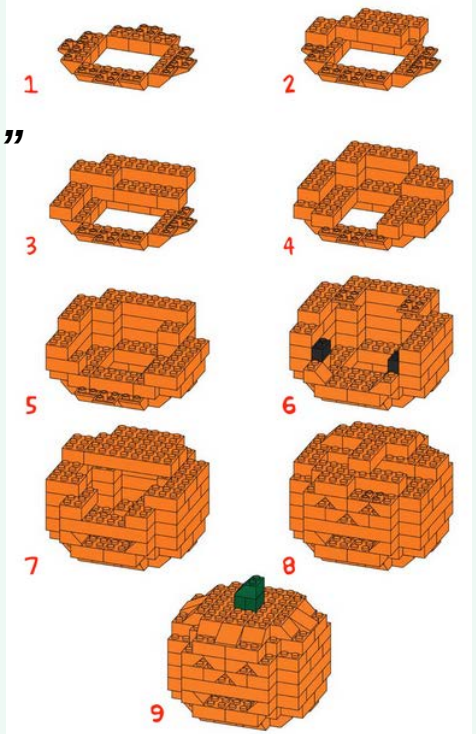
N. Przulj, D. G. Corneil, and I. Jurisica, "Modeling Interactome: Scale Free or Geometric?," *Bioinformatics*, vol. 20, num. 18, pg. 3508-3515, 2004.

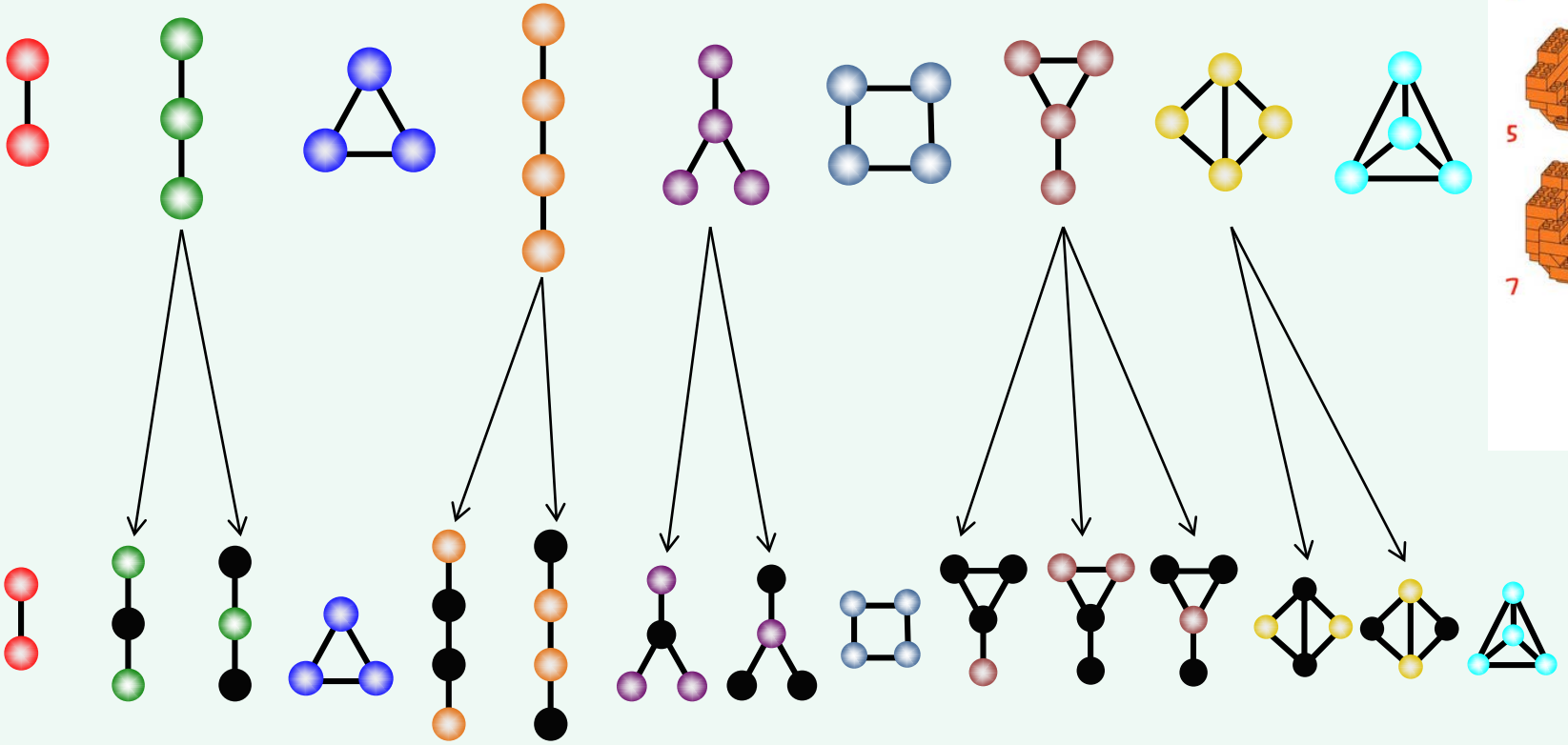
2. Novel Methods
Mine the Medical World of Inter-Connected Entities

ERC StG: 278212 (2012-2017): "Biological Network Topology Complements Genome as a Source of Biological Information"

Graphlets

"Legos of Networks"

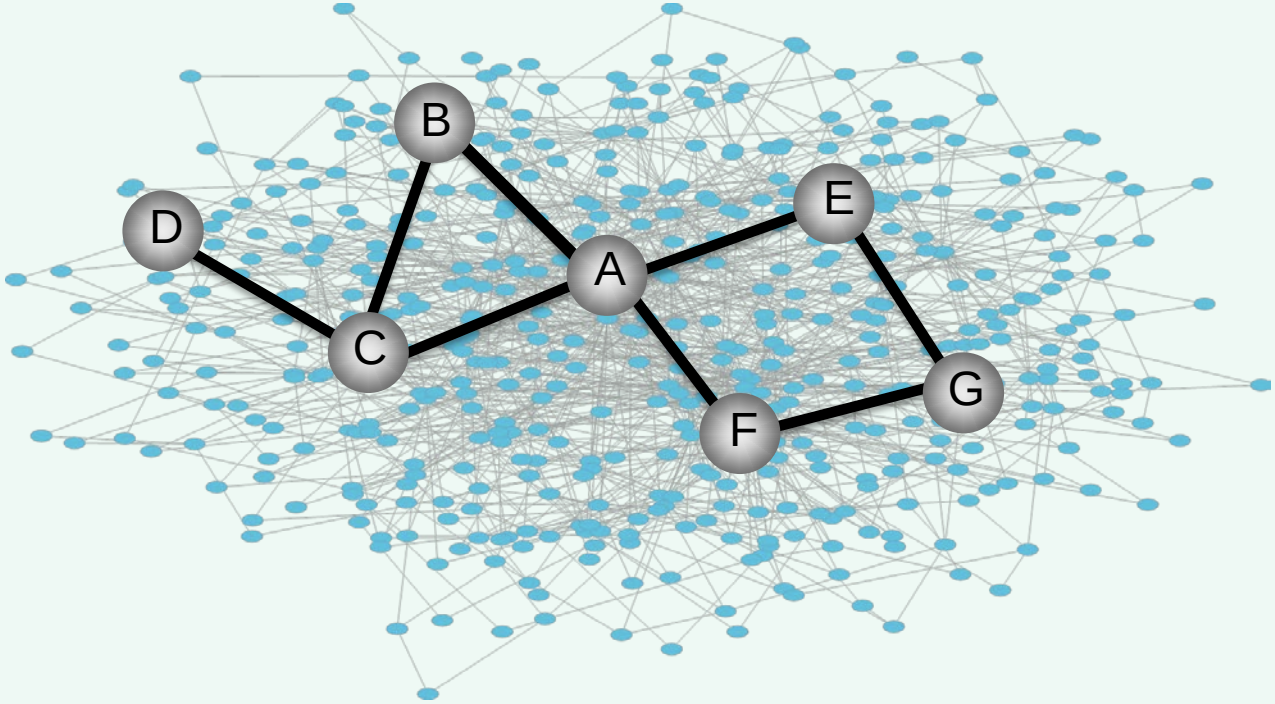




N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *Proceedings of the 2006 European Conference on Computational Biology, ECCB '06*, Eilat, Israel, January 21-24, 2007, **acceptance rate 18%.** *Bioinformatics*, volume 23, pages e177-e183, 2007

2. Novel Methods

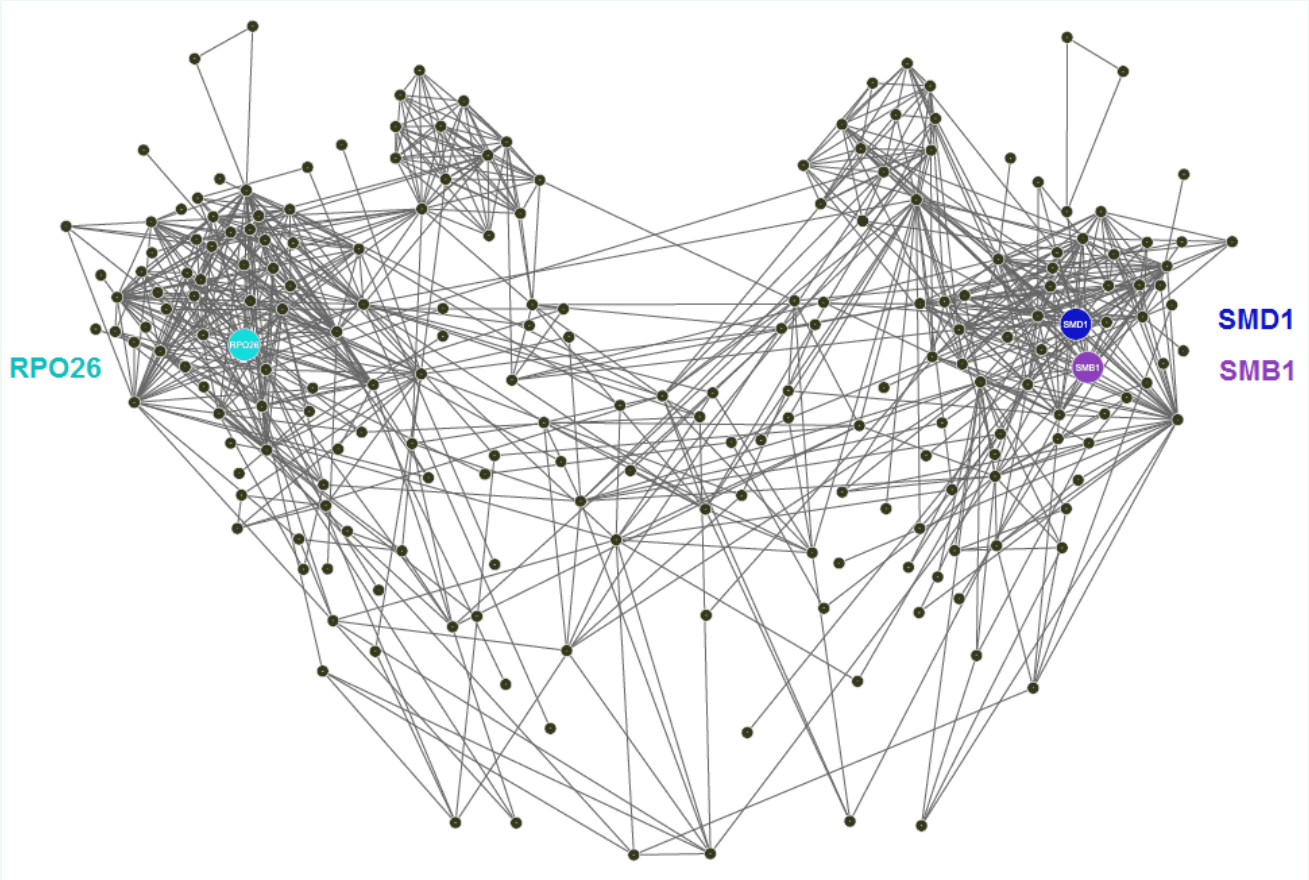
Mine the Medical World of Inter-Connected Entities



Orbit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	4	3	5	1	0	6	0	2	1	0	1	2	0	0	0
B	2	3	0	1	2	0	1	0	0	0	3	0	0	0	0
C	3	2	2	1	2	2	1	0	0	0	2	1	0	0	0
D	1	2	0	0	2	0	0	0	0	1	0	0	0	0	0
E	2	4	1	0	1	2	2	0	1	1	0	0	0	0	0
F	2	4	1	0	1	2	2	0	1	1	0	0	0	0	0
G	2	2	1	0	4	0	0	0	1	0	0	0	0	0	0

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

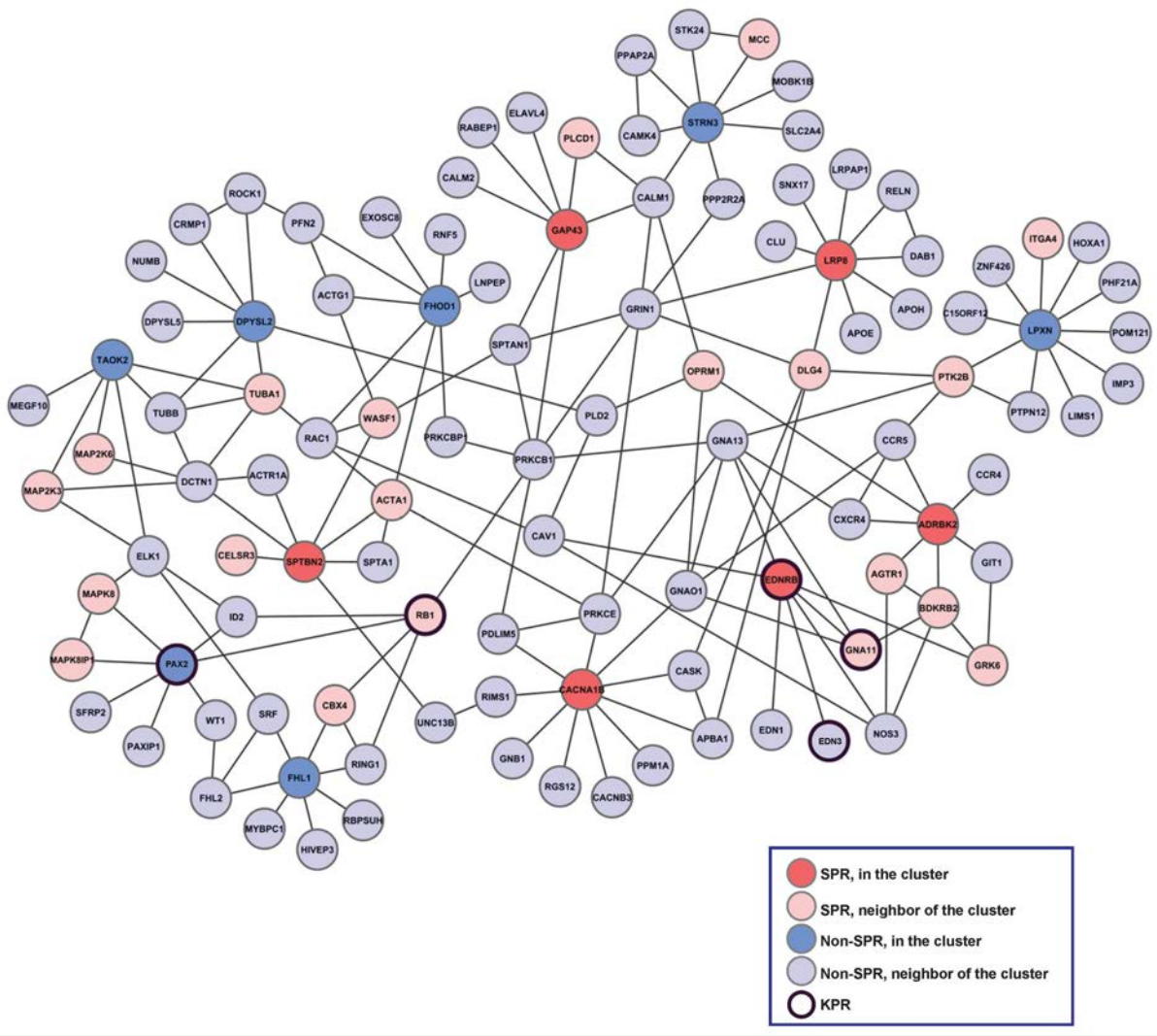
90% similar wiring – significantly enriched:



→ Biological function
→ Protein complexes
→ Sub-cellular localization
→ Tissue expression
→ Disease

T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures", *Cancer Informatics*, vol. 4, pg. 257-273, 2008 (Highly accessed)

2. Novel Methods
Mine the Medical World of Inter-Connected Entities



Cancer research:

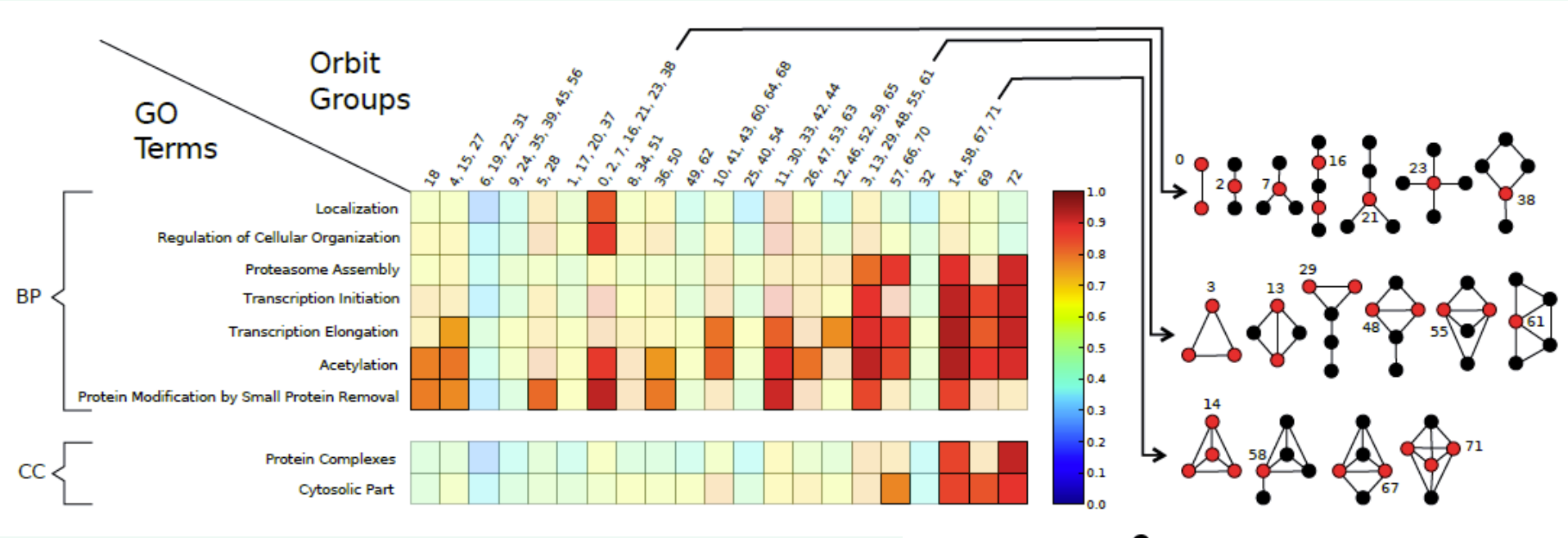
→ New proteins for melanin production

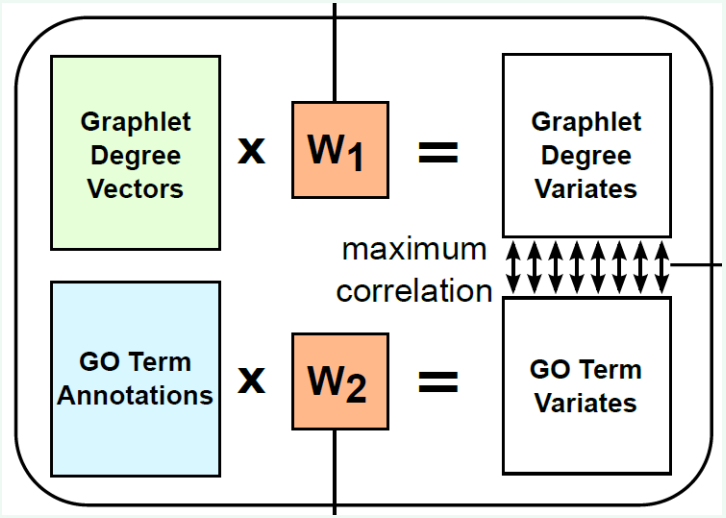
→ Same cancer type: more similar wiring

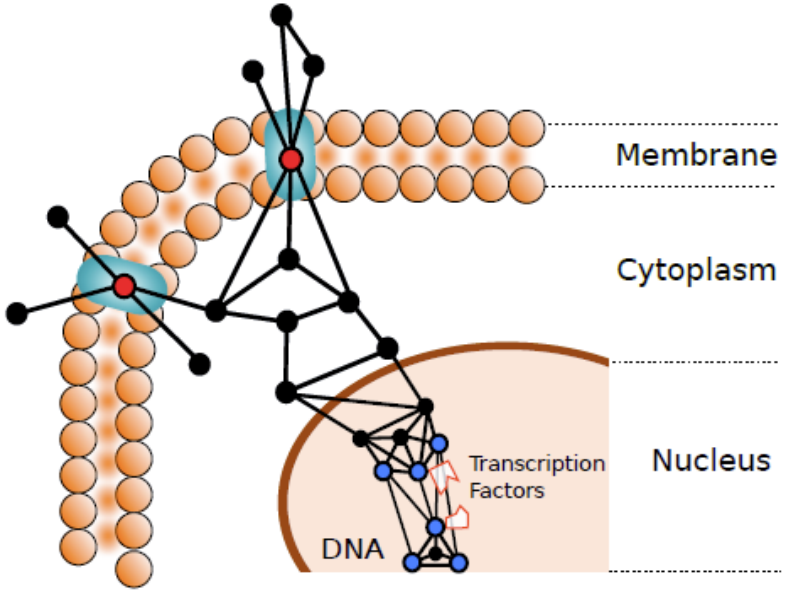
→ Far away in the network

T. Milenković, V. Memisević, A. K. Ganesan, and **N. Pržulj**, *J. Roy. Soc. Interface*, 7(44):423-437, 2010
H. Ho, T. Milenković, V. Memisević, J. Aruri, **N. Pržulj**, and A. K. Ganesan, *BMC Systems Biology,* 4:84, 2010 (Highly accessed)

2. Novel Methods
Mine the Medical World of Inter-Connected Entities



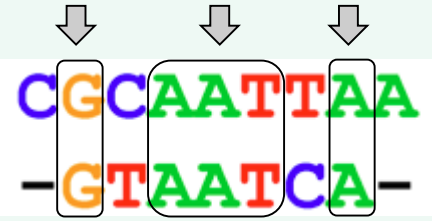


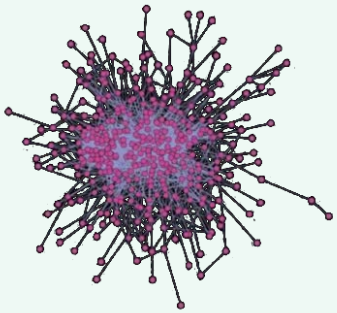


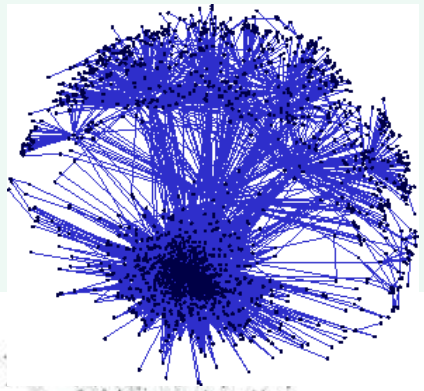
D. Davis, O. N. Yaveroglou, N. Malod-Dognin, A. Stojmirovic, **N. Przulj**, "Topology-Function Conservation in Protein-Protein Interaction Networks," *Bioinformatics* 31(10):1632-1639, 2015. IF=7.3

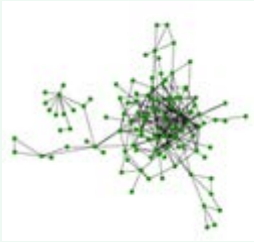
2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Network Alignment



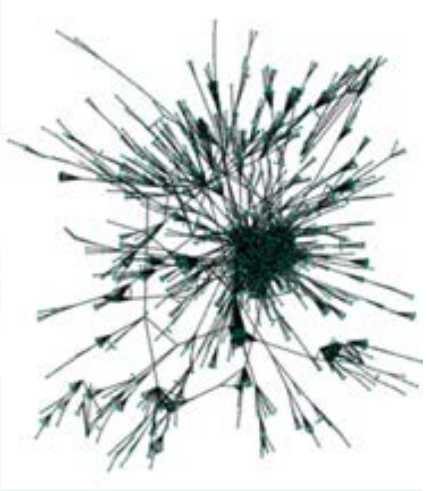




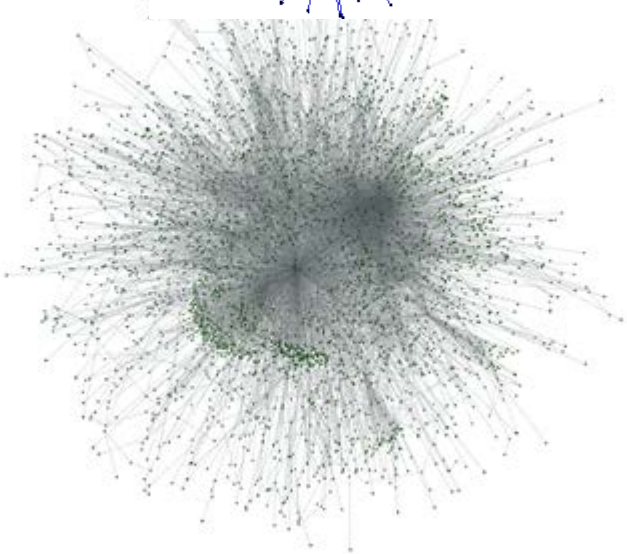


Isorank:
116 nodes
261 edges

GRAAL:
267 nodes
900 edges



MI-GRAAL:
1,858 nodes
3,467 edges



L-GRAAL:
5,726 nodes
16,084 edges
Yeast: 98% proteins
21% interactions

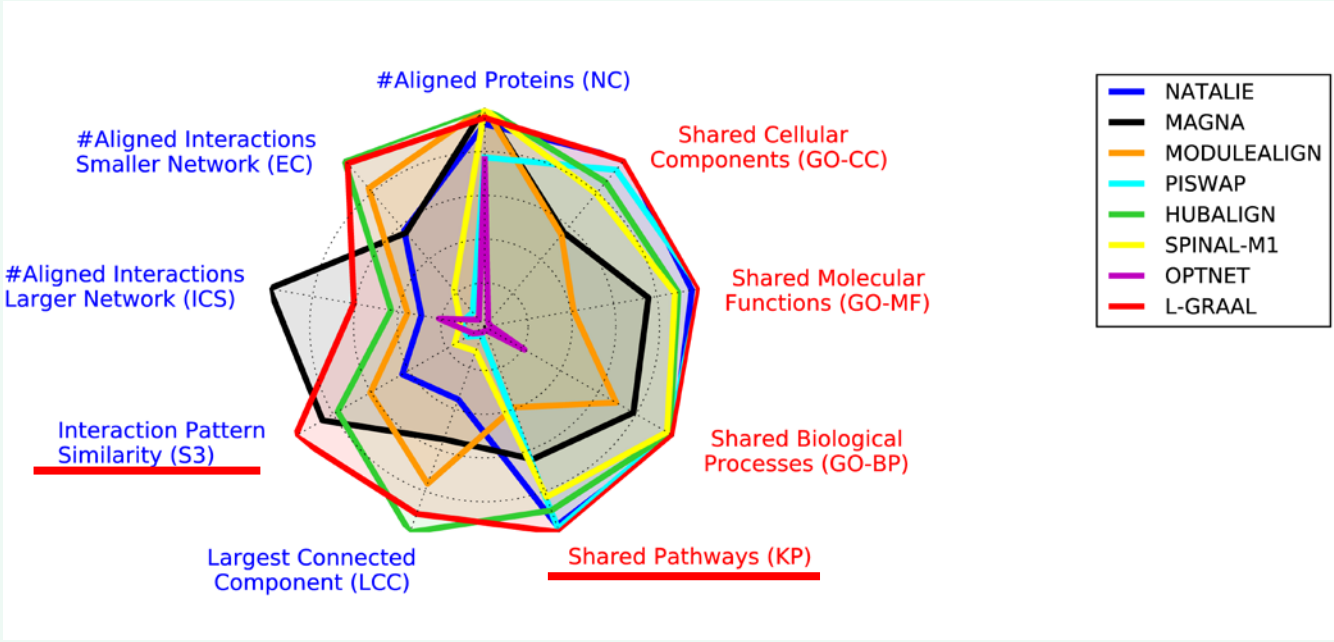
N. Malod-Dognin & N. Pržulj, L-GRAAL, *Bioinformatics*, doi: 10.1093/bioinformatics/btv130, 2015
N. Malod-Dognin & N. Pržulj, GR-ALIGN, *Bioinformatics*, doi:10.1093/bioinformatics/btu020, 2014
V. Memisevic & N. Pržulj, C-GRAAL, *Integrative Biology*, doi:10.1039/c2ib00140c, 2012
O. Kuchaiev **&** N. Pržulj, MI-GRAAL, *Bioinformatics*, 27(10): 1390-6, 2011
O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes, & N. Pržulj**,** *J. Royal Society Interface,* 7:1341-1354, 2010
T. Milenkovic, W.L. Wong, W. Hayes, & N. Pržulj, *Cancer Informatics*, 9:121-37, June 30, 2010 (Highly visible)

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Alignment of PPI Networks – ***Ulign***

- Many methods
- All heuristic
- No gold standard

- Questions:
  - Which aligner for which data?
  - Which scoring scheme for evaluation?
  - Coverage: biological and topological?
  - Contribution of topology vs sequence?



- Map biologically and topologically ***different*** network regions
- Each covers only about 50% of the proteins of the larger network

- **Together** – map **entire** networks → ***Ulign***
  - Biologically coherent

- The most topologically coherent – using topology only
- The most biologically coherent – using sequence only

Why?
- Existing annotations ill-suited?
- Methodological limitations?

→ **Combine** topology and sequence information

N. Malod-Dognin, K. Ban and **N. Przulj,** Unified Alignment of Protein-Protein Interaction Networks, *Scientific Reports- Nature*, 7:953, 2017

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

- ✓ The best performing
- ✓ Robust
- ✓ …

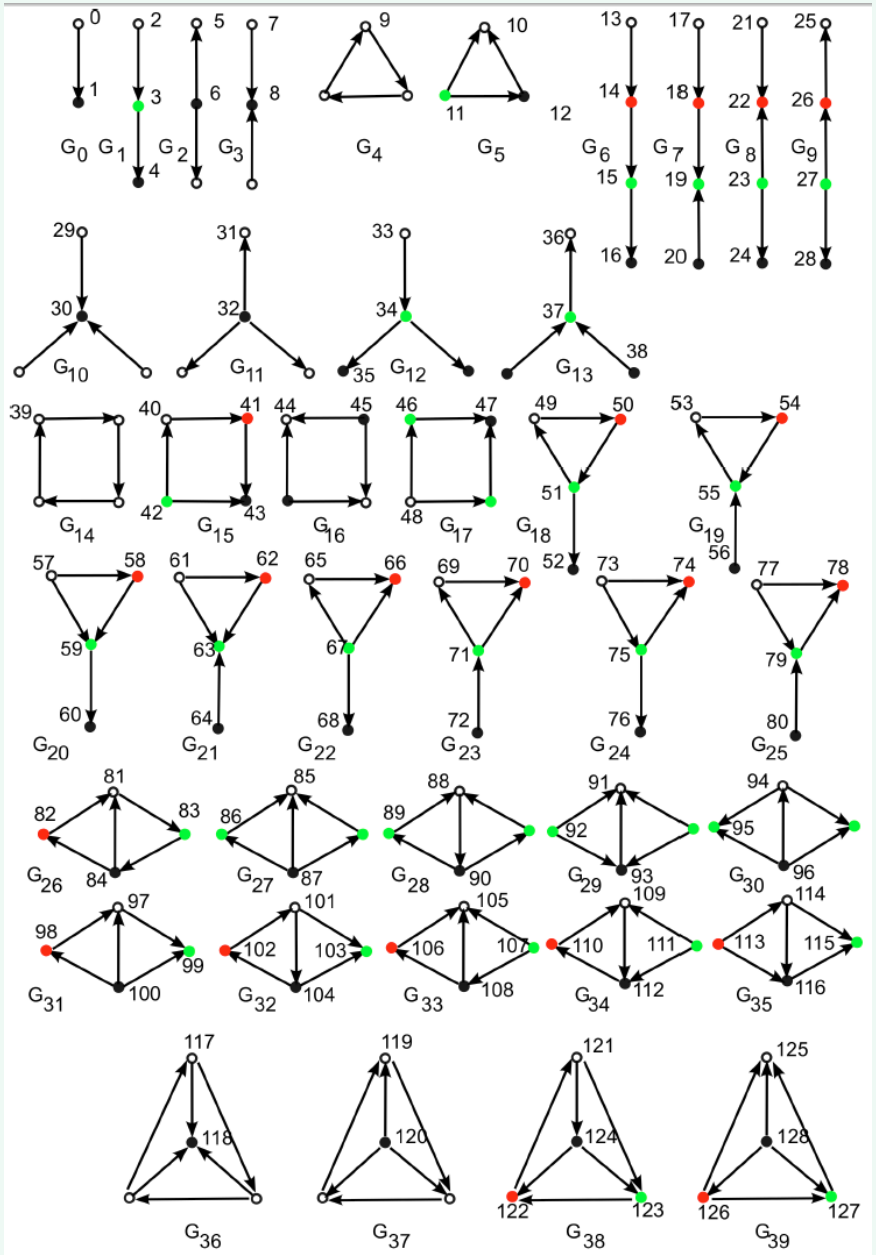
- ❑ PPI networks are ***geometric***

N. Przulj, D. G. Corneil, and I. Jurisica, "Modeling Interactome: Scale Free or Geometric?," *Bioinformatics*, vol. 20, num. 18, pg. 3508-3515, 2004.

N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," Proceedings of the 2006 European Conference on Computational Biology, ECCB '06, Eilat, Israel, January 21-24, 2007, **acceptance rate 18%.** *Bioinformatics*, volume 23, pages e177-e183, 2007

…

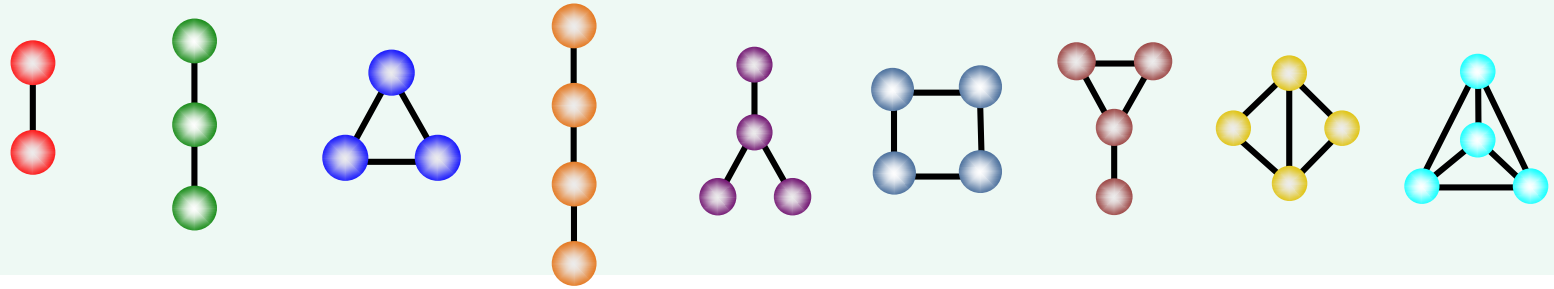
- ➢ **Directed Networks**
- ➢ **Track dynamics**

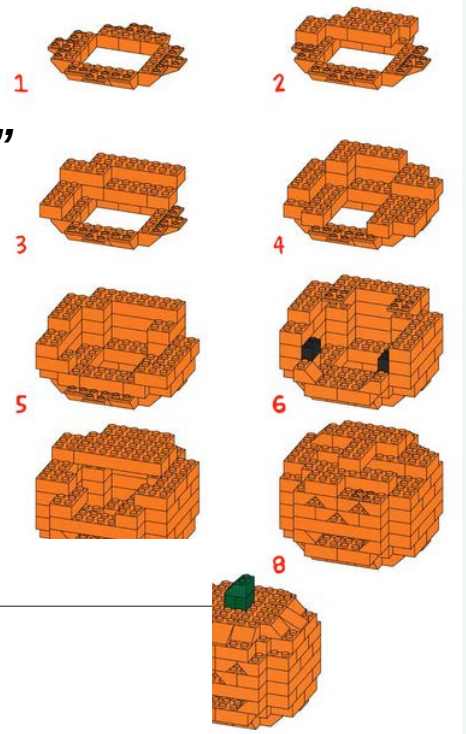


A. Sarajlic, N. Malod-Dognin, O. N. Yaveroglou, and **N. Przulj,** "Graphlet-based Characterization of Directed Networks," *Scientific Reports - Nature*, 6:35098, 2016

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Graphlets
"Legos of Networks"





INSIGHTS | *PERSPECTIVES*

Network analytics in the age of Big Data

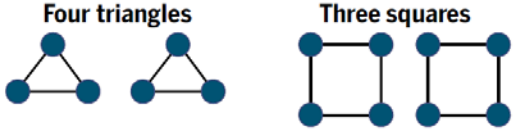
How can we holistically mine big data?

By **Nataša Pržulj** *and* **Noël Malod-Dognin**

We live in a complex world of interconnected entities. In all areas of human endeavor, from biology to medicine, economics, and climate science, we are flooded with large-scale data sets. They describe intricate real-world systems from different and complementary viewpoints, with entities being modeled as nodes and their connections as edges, comprising large networks. This is into RNAs and translated into proteins, which adopt various three-dimensional structures to carry out particular cellular functions. Molecular interactions are captured by different high-throughput biotechnologies and modeled with different types of networks. Individual analyses of molecular networks have revealed that molecules involved in similar functions tend to group together in a network and are similarly wired (*13*), leading to better understanding of gene functions (*6*) and molecular organization of the cell (*7*) and to im-

Network structures

The four networks shown have exactly the same size (the same number of nodes and edges), and each node within each network has the same degree (the number of interactions with other nodes), but each network canis of very different structure.

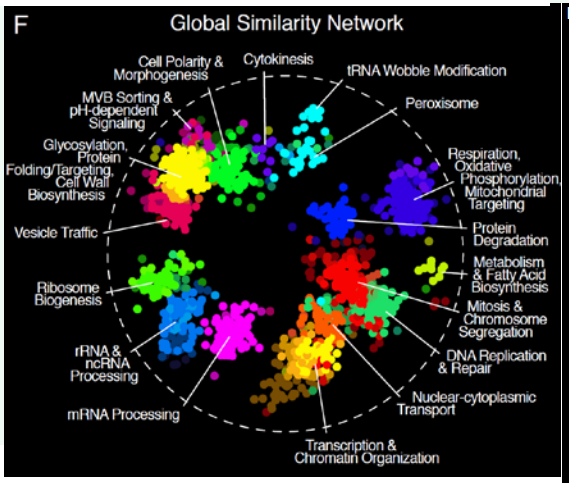


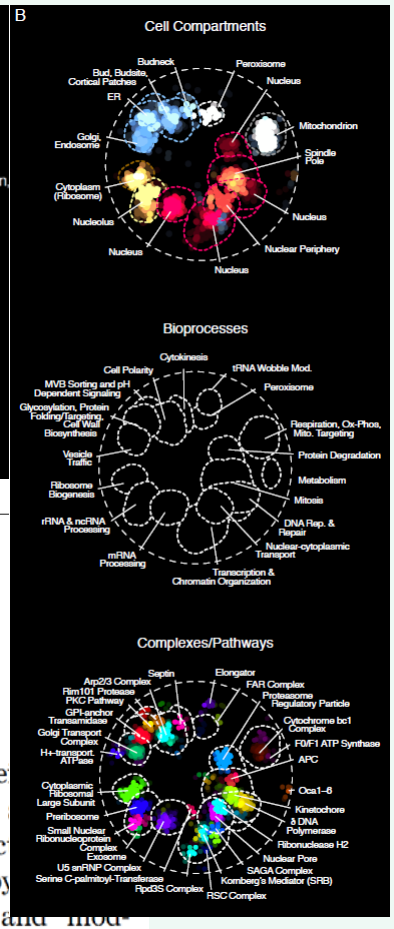
2. Novel Methods
Mine the Medical World of Inter-Connected Entities

A global genetic interaction network maps a wiring diagram of cellular function

Science 23 Sep 2016:
Vol. 353, Issue 6306, aaf1420
DOI: 10.1126/science.aaf1420

Michael Costanzo[1*], Benjamin VanderSluis[2,3*], Elizabeth N. Koch[2*], Anastasia Baryshnikova[4*], Carles Pons[2*≈], Guihong Tan[1*], Wen Wang[2], Matej Usaj[1], Julia Hanchard[1,5], Susan D. Lee[6], Vincent Pelechano[7◊], Erin B. Styles[1,5], Maximilian Billmann[8], Jolanda van Leeuwen[1], Nydia van Dyk[1], Zhen-Yuan Lin[9], Elena Kuzmin[1,5], Justin Nelson[2,10], Jeff S. Piotrowski[1,11§], Tharan Srikumar[12], Sondra Bahr[1], Yiqun Chen[1], Raamesh Deshpande[2], Christoph F. Kurat[1¶], Sheena C. Li[1,11], Zhijian Li[1], Mojca Mattiazzi Usaj[1], Hiroki Okada[13], Natasha Pascoe[1,5], Bryan-Joseph San Luis[1], Sara Sharifpoor[1], Emira Shuteriqi[1], Scott W. Simpkins[2,10], Jamie Snider[1], Harsha Garadi Suresh[1], Yizhao Tan[1], Hongwei Zhu[1], Noel Malod-Dognin[14], Vuk Janji[14], Natasa Przulj[14,15], Olga G. Troyanskaya[3,4], Igor Stagljar[1,5,16], Tian Xia[2,17], Yoshikazu Ohya[13], Anne-Claude Gingras[5,9], Brian Raught[12], Michael Boutros[8], Lars M. Steinmetz[7,18], Claire L. Moore[6], Adam P. Rosebrock[1,5], Amy A. Caudy[1,5], Chad L. Myers[2,10#], Brenda Andrews[1,5#], and Charles Boone[1,5,#]





Network analytics in the age of Big Data

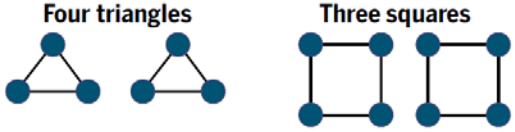
How can we holistically mine big data?

By **Nataša Pržulj** *and* **Noël Malod-Dognin**

We live in a complex world of interconnected entities. In all areas of human endeavor, from biology to medicine, economics, and climate science, we are flooded with large-scale data sets. They describe intricate real-world systems from different and complementary viewpoints, with entities being modeled as nodes and their connections as edges, comprising large networks. This is

Network structures

The four networks shown have exactly the same size (the same number of nodes and edges), and each node within each network has the same degree (the number of interactions with other nodes), but each network canis of very different structure.



into RNAs and translated into prote
adopt various three-dimensional
to carry out particular cellular funct
lecular interactions are captured by
high-throughput biotechnologies and mod-
eled with different types of networks. Individual analyses of molecular networks have revealed that molecules involved in similar functions tend to group together in a network and are similarly wired (*13*), leading to better understanding of gene functions (*6*) and molecular organization of the cell (*7*) and to im-

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Published online: March 15, 2017

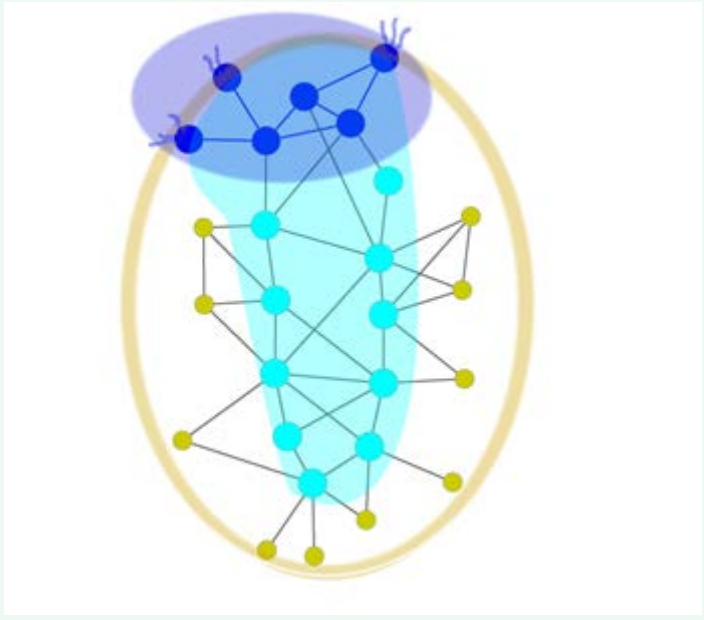
Article

TRANSPARENT PROCESS OPEN ACCESS

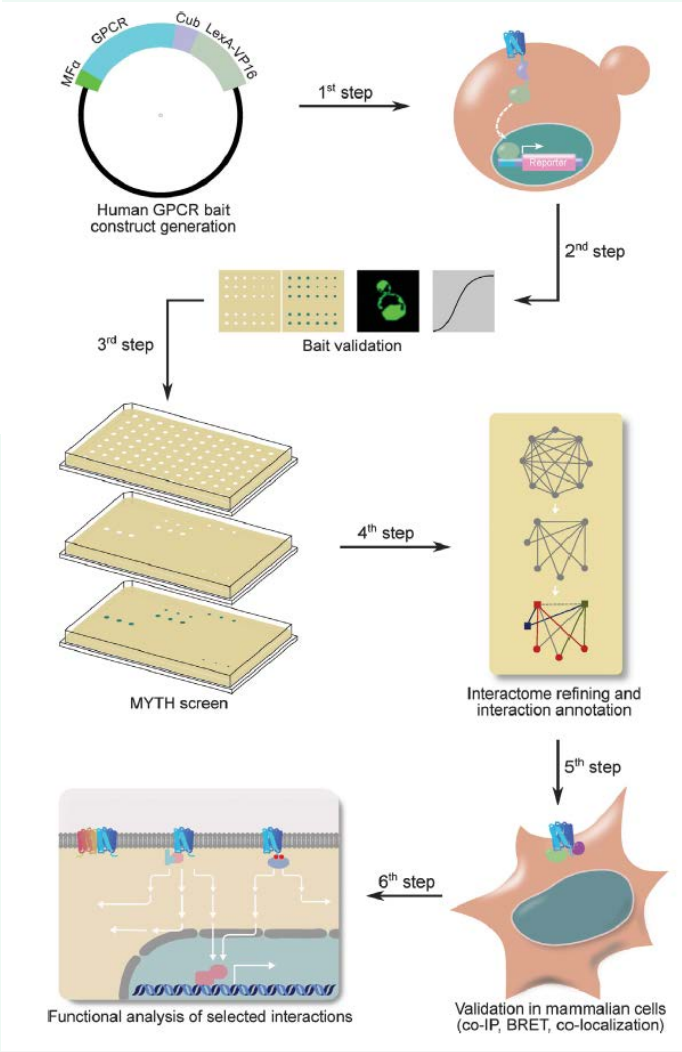
molecular systems biology

Systematic protein–protein interaction mapping for clinically relevant human GPCRs

Kate Sokolina[1,†], Saranya Kittanakom[1,†], Jamie Snider[1,†], Max Kotlyar[2], Pascal Maurice[3,4,5,6], Jorge Gandía[7,8], Abla Benleulmi-Chaachoua[3,4,5], Kenjiro Tadagaki[3,4,5], Atsuro Oishi[3,4,5], Victoria Wong[1], Ramy H Malty[9], Viktor Deineko[9], Hiroyuki Aoki[9], Shahreen Amin[9], Zhong Yao[1], Xavier Morató[7,8], David Otasek[2], Hiroyuki Kobayashi[10], Javier Menendez[1], Daniel Auerbach[11], Stephane Angers[12], Natasa Pržulj[13], Michel Bouvier[10], Mohan Babu[9], Francisco Ciruela[7,8], Ralf Jockers[3,4,5], Igor Jurisica[2,14,15] & Igor Stagljar[1,16,17,*]



- ✓ "Spine" of the network
  - "Dominating set" heuristic
- ✓ Functionally and topologically separates the cell
- ✓ Predict new GPCRs:
  - e.g., chromosome 20 open reading frame 39 (TMEM90B)

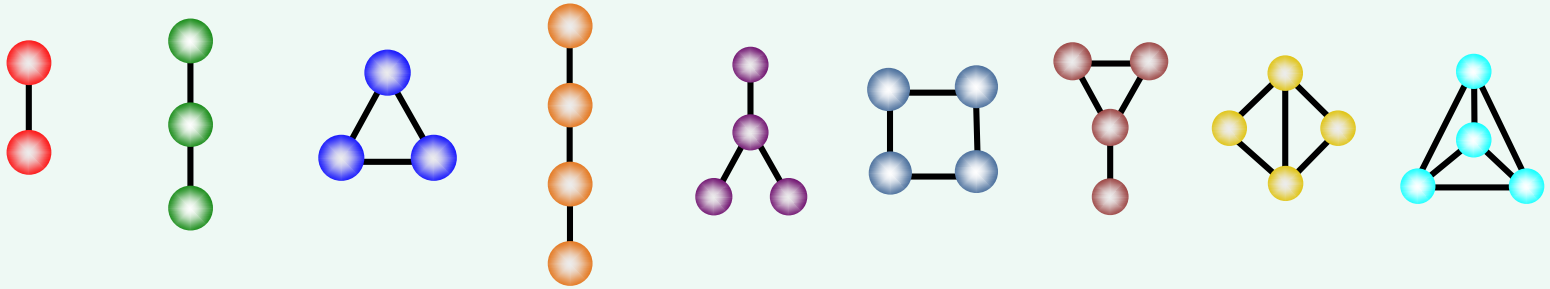


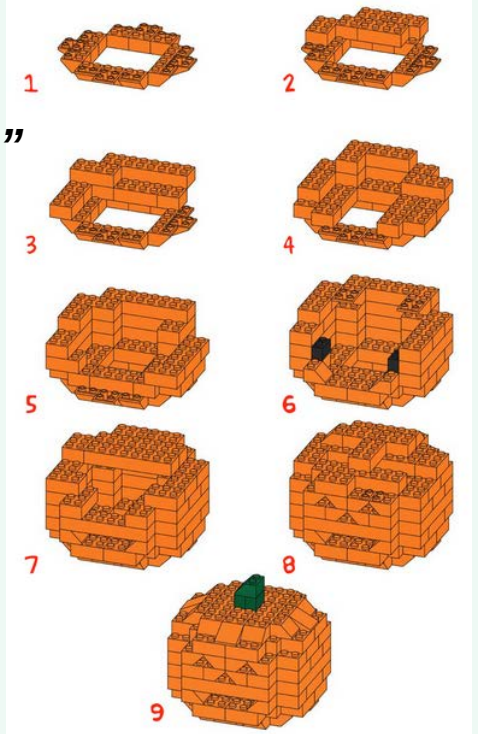
Workflow for generating the human full-length GPCR interactome.

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Graphlets

"Legos of Networks"





https://scholar.google.ca/scholar?hl=en&q=graphlets&btnG=&as_sdt= 90% Search

Web Images More... Sign in

Google graphlets

Scholar About 1,490 results (0.05 sec) My Citations

Articles
Case law
My library

Any time
Since 2017
Since 2016
Since 2013
Custom range...

Sort by relevance
Sort by date

include patents
include citations

Discovering discriminative graphlets for aerial image categories recognition
L Zhang, Y Han, Y Yang, M Song... - IEEE Transactions on ..., 2013 - ieeexplore.ieee.org
Abstract: Recognizing aerial image categories is useful for scene annotation and surveillance. Local features have been demonstrated to be robust to image transformations, including occlusions and clutters. However, the geometric property of an aerial image (ie,
Cited by 111 Related articles All 10 versions Cite Save

Probabilistic graphlet transfer for photo cropping
L Zhang, M Song, Q Zhao, X Liu, J Bu... - IEEE Transactions on ..., 2013 - ieeexplore.ieee.org
Abstract: As one of the most basic photo manipulation processes, photo cropping is widely used in the printing, graphic design, and photography industries. In this paper, we introduce graphlets (ie, small connected subgraphs) to represent a photo's aesthetic features, and
Cited by 95 Related articles All 16 versions Cite Save

Model selection for social networks using graphlets
J Janssen, M Hurshman, N Kalyaniwalla - Internet Mathematics, 2012 - Taylor & Francis
Several network models have been proposed to explain the link structure observed in online social networks. This paper addresses the problem of choosing the model that best fits a given real-world network. We implement a model-selection method based on unsupervised
Cited by 25 Related articles All 10 versions Cite Save

Integrating local features into discriminative graphlets for scene classification
L Zhang, W Bian, M Song, D Tao, X Liu - International Conference on ..., 2011 - Springer
Abstract Scene classification plays an important role in multimedia information retrieval. Since local features are robust to image transformation, they have been used extensively for scene classification. However, it is difficult to encode the spatial relations of local features in
Cited by 9 Related articles All 4 versions Cite Save

Ego-centric graphlets for personality and affective states recognition
S Teso, J Staiano, B Lepri, A Passerini... - Social Computing (..., 2013 - ieeexplore.ieee.org
Abstract: Do we tend to perceive ourselves more creative when surrounded by creative people? Or rather the opposite holds? Such information is very valuable to understand how to optimize work processes and boost people's productivity along with their happiness and
Cited by 5 Related articles All 14 versions Cite Save

From quasirandom graphs to graph limits and graphlets
F Chung - Advances in Applied Mathematics, 2014 - Elsevier
Abstract We generalize the notion of quasirandomness which concerns a class of equivalent properties that random graphs satisfy. We show that the convergence of a graph sequence under the spectral distance is equivalent to the convergence using the (normalized) cut
Cited by 4 Related articles All 10 versions Cite Save

Overview

Medicine: complex world of inter-connected entities

1. **Motivation**
2. **New Methods – Examples:** mine inter-connected data
   i. **Single layer of omics data: Molecular networks** → function, disease
   ii. **Multiple layers of heterogeneous data:**
      - **Patient-centered data integration → Precision medicine**
      - **Disease re-classification**
      - **Gene Ontology reconstruction**
      - **Network alignment**
3. **Vision**

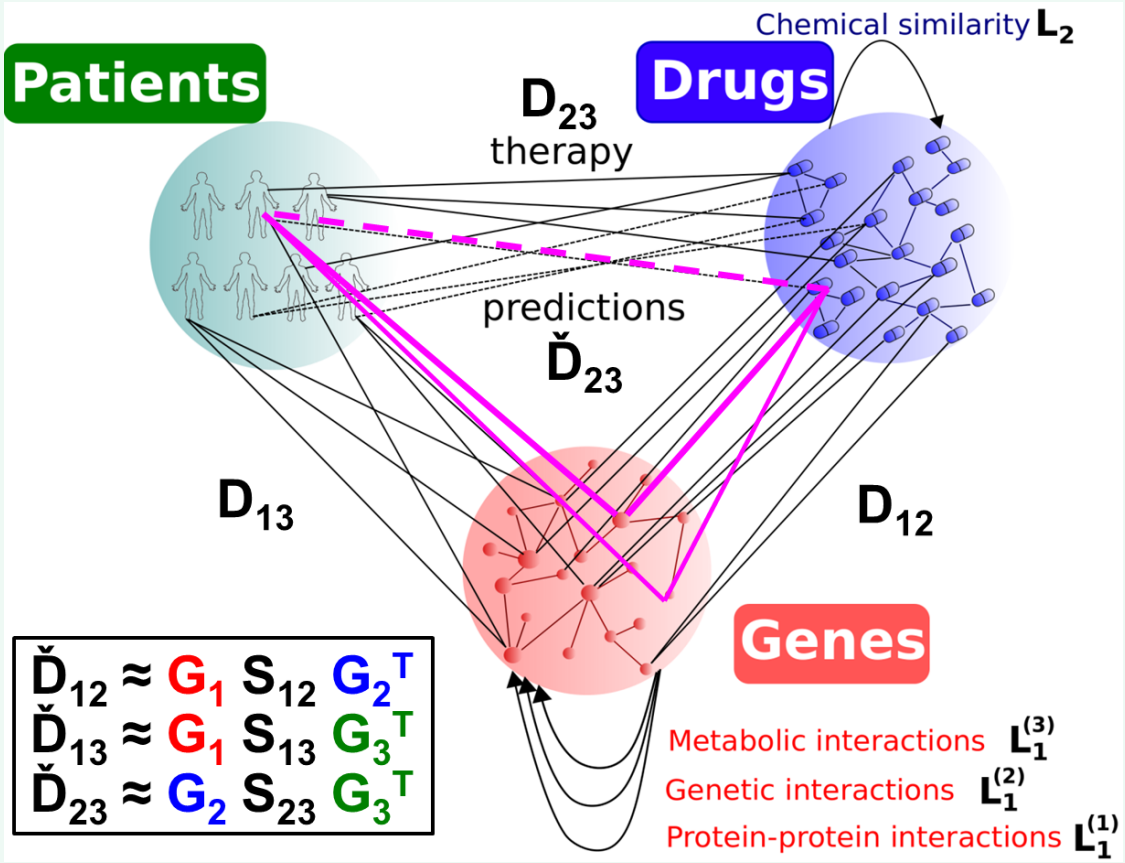
2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Multi-disciplinary, data-fusion methodology

Motivation:

- Captures all systems-level
- Captures how data relate
- Mechanistic explanations

PNMTF → Personalize Treatment



$$\min\{\sum_{1\le i\le j\le p} [\, ||W_{ij} \circ (D_{ij} - G_i S_{ij} G_j^T)||^2 + \alpha\, ||S_{ij}||^2 + \alpha_i\, tr(G_i^T L_i G_i) + \alpha_j\, tr(G_j^T L_j G_j)] : G_i, S_{ij} \ge 0\}$$

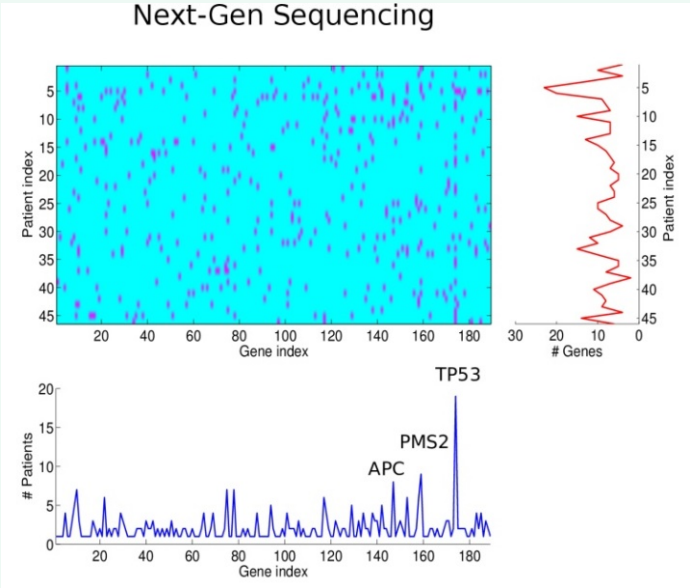
$\alpha\, ||S_{ij}||^2$ maintain sparsity of S_{ij} , $\alpha_i\, tr(G_i^T L_i G_i)$ and $\alpha_j\, tr(G_j^T L_j G_j)$ adding prior knowledge (penalties), G_i, $S_{ij} \ge 0$ is needed for cluster interpretation

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

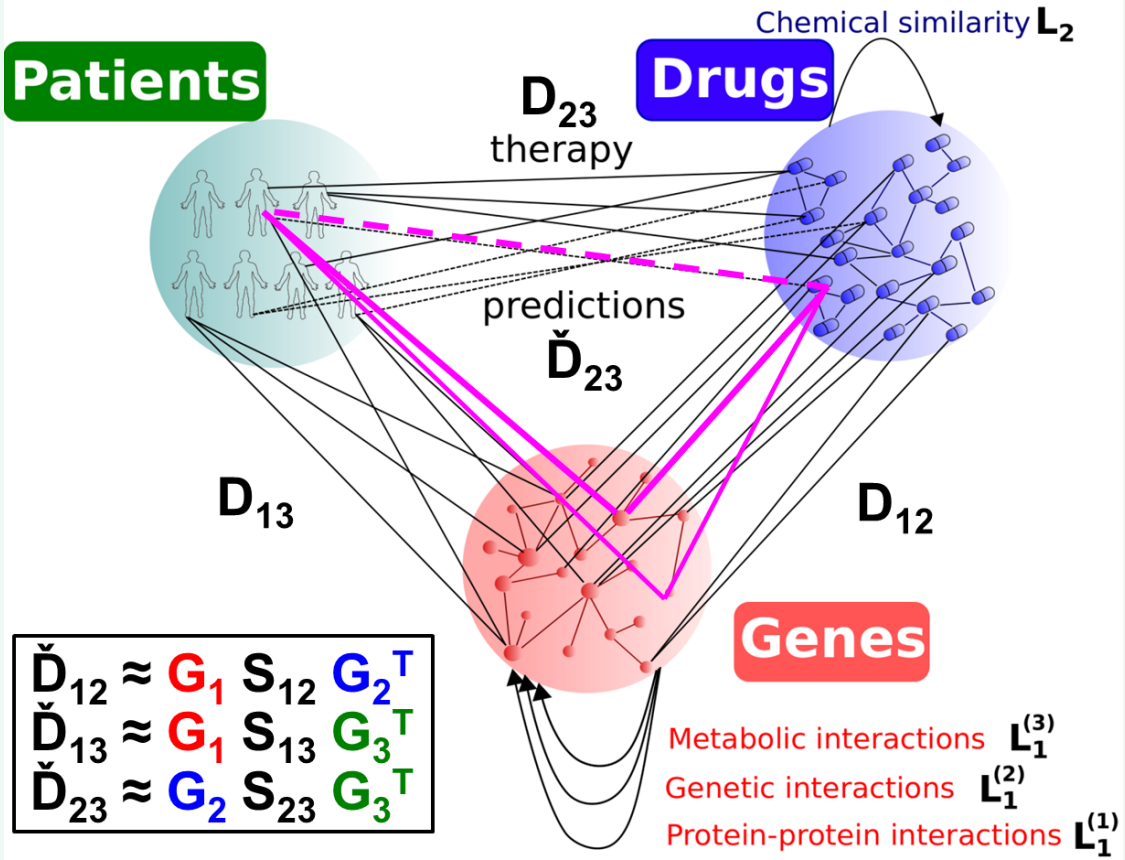
Multi-disciplinary, data-fusion methodology

Motivation:

- Captures all systems-level
- Captures how data relate
- Mechanistic explanations



PNMTF → Personalize Treatment



$$\min\{\sum_{1\le i\le j\le p} [\, \|W_{ij} \circ (D_{ij} - G_i S_{ij} G_j^T)\|^2 + \alpha \|S_{ij}\|^2 + \alpha_i \, tr(G_i^T L_i G_i) + \alpha_j \, tr(G_j^T L_j G_j)] : G_i, S_{ij} \ge 0\}$$

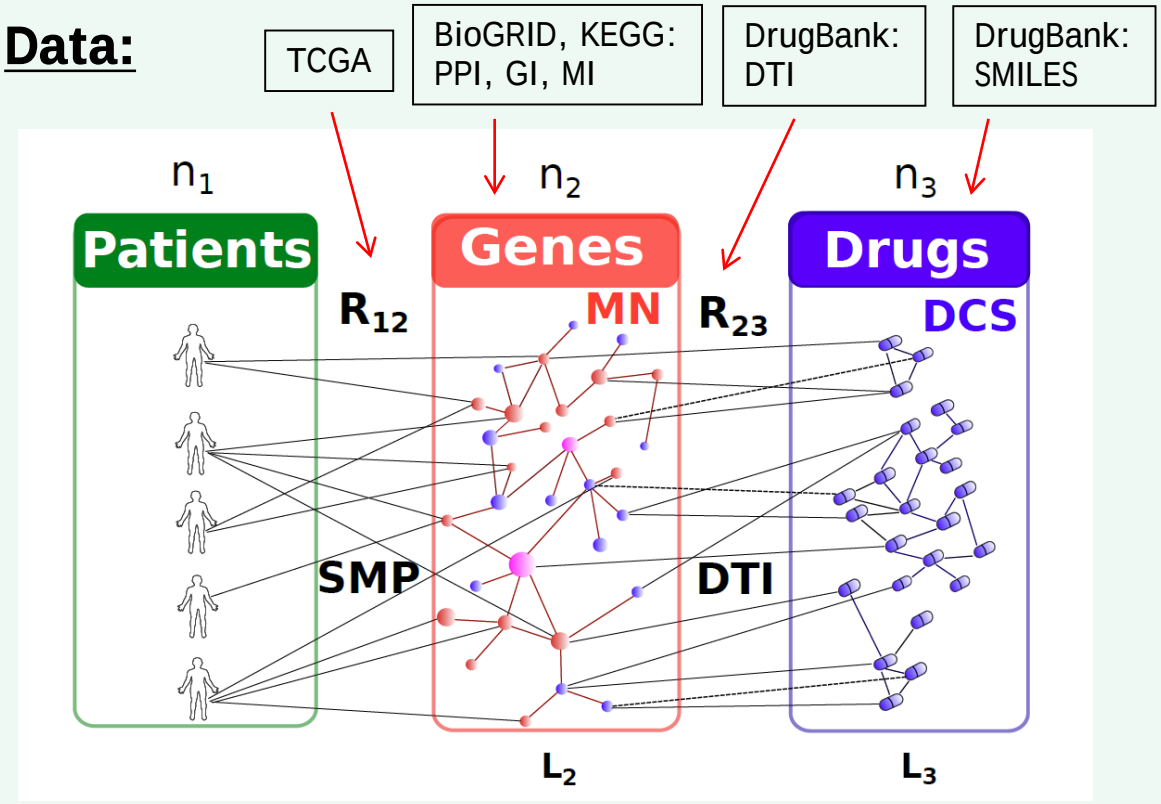
$\alpha \|S_{ij}\|^2$ maintain sparsity of S_{ij} , $\alpha_i \, tr(G_i^T L_i G_i)$ and $\alpha_j \, tr(G_j^T L_j G_j)$ adding prior knowledge (penalties), G_i, $S_{ij} \ge 0$ is needed for cluster interpretation

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Patient-Specific Data Fusion → Personalized Treatment

Co-clustering: patients, genes and drugs

Data:



353 serous ovarian cancer patients from TCGA:

1. Patient stratification
2. Driver gene prediction
3. Drug repurposing

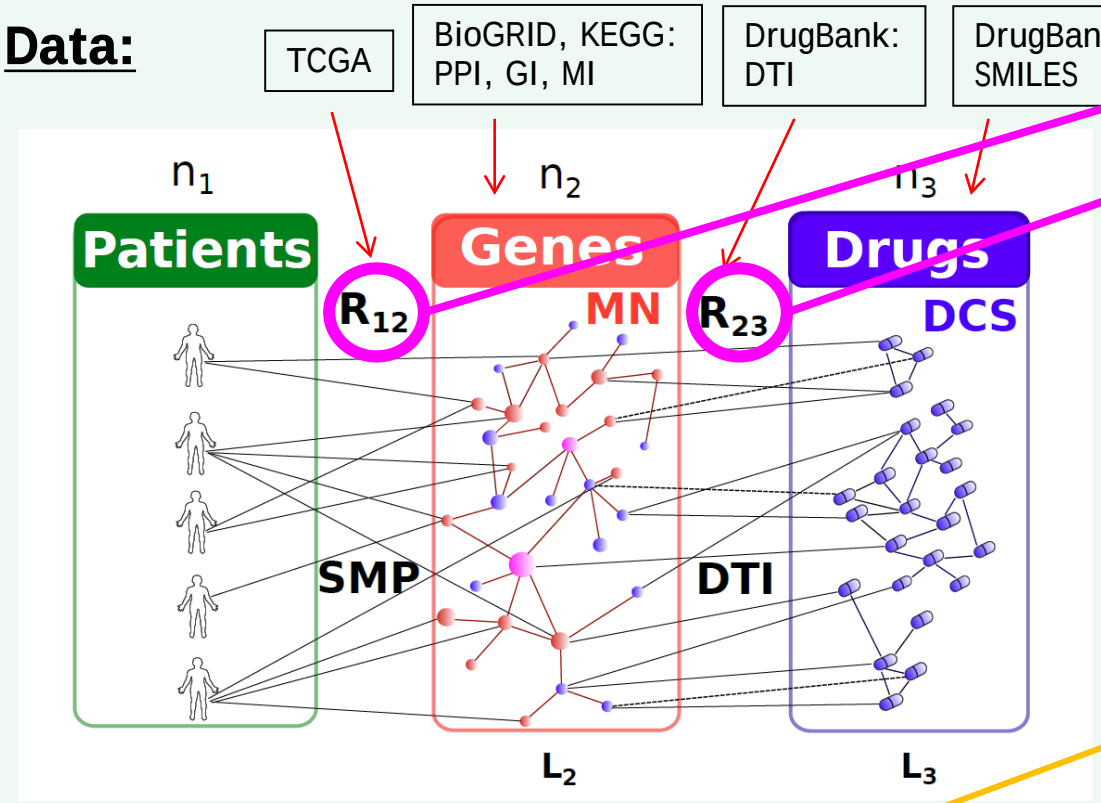
V. Gligorijevic, N. Malod-Dognin and **N. Przulj**, Patient-specific data fusion for cancer stratification and personalized treatment, *PSB*, 2016

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Patient-Specific Data Fusion → Personalized Treatment

Co-clustering: patients, genes and drugs

Data:



$$R_{12} \approx G_1 H_{12} G_2^T$$
$$R_{23} \approx G_2 H_{23} G_3^T$$

$k_1 \ll n_1$ – patient clusters
$k_2 \ll n_2$ – gene clusters
$k_3 \ll n_3$ – drug clusters
G_1, G_2 and G_3 are cluster indicator matrices

Ovarian cancer patients:

1. Patient stratification → $\hat{C}_1$
2. Driver gene prediction → $\hat{C}_2$
3. Drug repurposing → $\hat{R}_{23}$

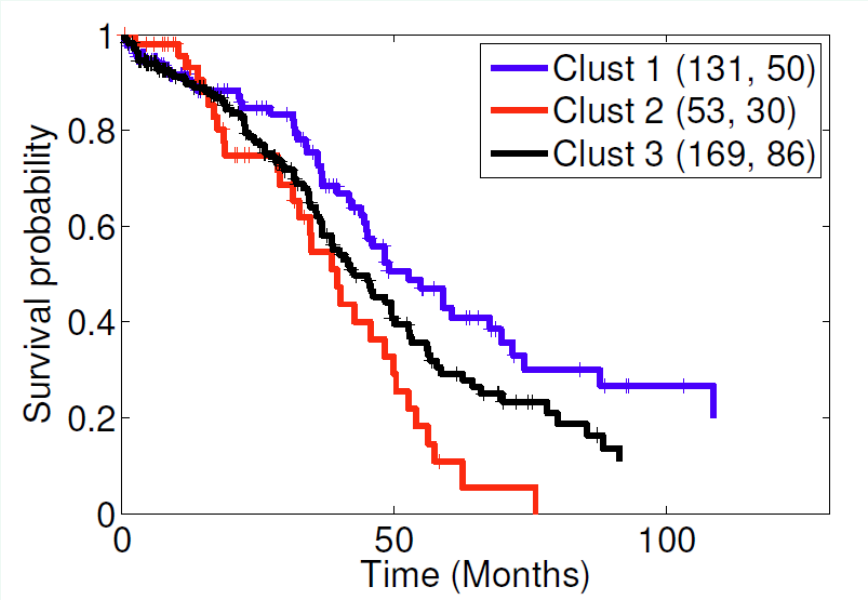
$$\min_{G_i \geq 0, 1 \leq i \leq 3} J = \min_{G_i \geq 0, 1 \leq i \leq 3} \left[\| R_{12} - G_1 H_{12} G_2^T \|_F^2 + \| R_{23} - G_2 H_{23} G_3^T \|_F^2 + tr(G_2^T L_2 G_2) + tr(G_3^T L_3 G_3) \right]$$

V. Gligorijevic, N. Malod-Dognin and **N. Przulj**, Patient-specific data fusion for cancer stratification and personalized treatment, *PSB*, 2016

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Patient-Specific Data Fusion → Personalized Treatment

Some results:



Kaplan-Meier survival curves for 3 patient groups found by GNMTF (log-rank p-val = 5.3×10^{-3})

$$\mathbf{R}_{12} \approx \mathbf{G}_1 \mathbf{H}_{12} \mathbf{G}_2^T$$
$$\mathbf{R}_{23} \approx \mathbf{G}_2 \mathbf{H}_{23} \mathbf{G}_3^T$$

$k_1 \ll n_1$ – patient clusters
$k_2 \ll n_2$ – gene clusters
$k_3 \ll n_3$ – drug clusters
$\mathbf{G}_1$, $\mathbf{G}_2$ and $\mathbf{G}_3$ are cluster indicator matrices

Ovarian cancer patients:

1. Patient stratification → $\hat{\mathbf{C}}_1$
2. Driver gene prediction → $\hat{\mathbf{C}}_2$
3. Drug repurposing → $\hat{\mathbf{R}}_{23}$

$$\min_{\mathbf{G}_i \geq 0, 1 \leq i \leq 3} J = \min_{\mathbf{G}_i \geq 0, 1 \leq i \leq 3} \Big[\| \mathbf{R}_{12} - \mathbf{G}_1 \mathbf{H}_{12} \mathbf{G}_2^T \|_F^2 + \| \mathbf{R}_{23} - \mathbf{G}_2 \mathbf{H}_{23} \mathbf{G}_3^T \|_F^2 + tr(\mathbf{G}_2^T \mathbf{L}_2 \mathbf{G}_2) + tr(\mathbf{G}_3^T \mathbf{L}_3 \mathbf{G}_3) \Big]$$

V. Gligorijevic, N. Malod-Dognin and **N. Przulj**, Patient-specific data fusion for cancer stratification and personalized treatment, *PSB*, 2016

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Patient-Specific Data Fusion → Personalized Treatment

Some results: ~40% of our 809 predicted driver genes in CCGD, Census, or IntOGen

New driver	Known drivers	Score	DB
ADAM32	BMPR2	1.000	–
REG1P	CLASP2	1.000	–
PCDHA2	CHD4	1.000	–
NCR1	BMPR2	1.000	–
USPL1	CLASP2	1.000	–
GDPD3	DDX5	1.000	–
LECT1	CLASP2	1.000	CCGD
IL25	CDK12, CCAR1	0.975	–
BAK1	ATRX, TFDP1, NDRG1	0.967	–
MOGAT2	ATRX, TFDP1, NDRG1	0.967	–
CHAF1A	ATRX, TFDP1, NDRG1	0.967	CCGD
PITX2	ATRX, TFDP1, NDRG1	0.967	–
SIN3B	ATRX, TFDP1, NDRG1	0.967	–
RPL30	ATRX, TFDP1, NDRG1	0.967	–
GRWD1	ATRX, TFDP1, NDRG1	0.967	–
SNAI1	ATRX, TFDP1, NDRG1	0.967	CCGD
RBMXP4	ATRX, TFDP1, NDRG1	0.967	–
CPNE7	ATRX, TFDP1, NDRG1	0.967	–
HIPK3	ATRX, TFDP1, NDRG1	0.967	CCGD
EPOR	ATRX, TFDP1, NDRG1	0.967	CCGD

↔ TGFs, cell proliferation & progression
↔ proliferation, migration, anti-apoptosis; prognosis markers

Ovarian cancer patients:

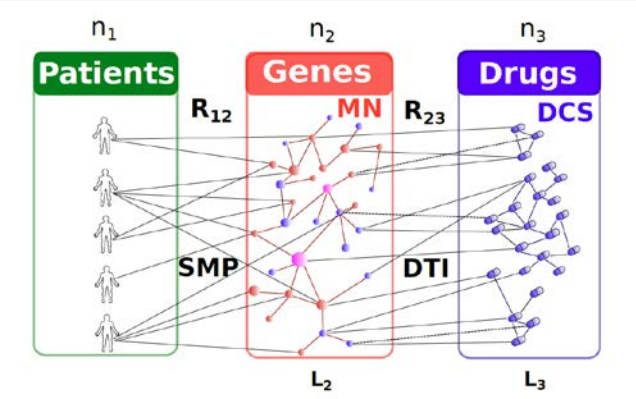
1. Patient stratification → $\hat{\mathbf{C}}_1$
2. **Driver gene prediction** → $\hat{\mathbf{C}}_2$
3. Drug repurposing → $\hat{\mathbf{R}}_{23}$

$$\min_{\mathbf{G}_i \geq 0, 1 \leq i \leq 3} J = \min_{\mathbf{G}_i \geq 0, 1 \leq i \leq 3} \Big[\| \mathbf{R}_{12} - \mathbf{G}_1 \mathbf{H}_{12} \mathbf{G}_2^T \|_F^2 + \| \mathbf{R}_{23} - \mathbf{G}_2 \mathbf{H}_{23} \mathbf{G}_3^T \|_F^2 + tr(\mathbf{G}_2^T \mathbf{L}_2 \mathbf{G}_2) + tr(\mathbf{G}_3^T \mathbf{L}_3 \mathbf{G}_3) \Big]$$

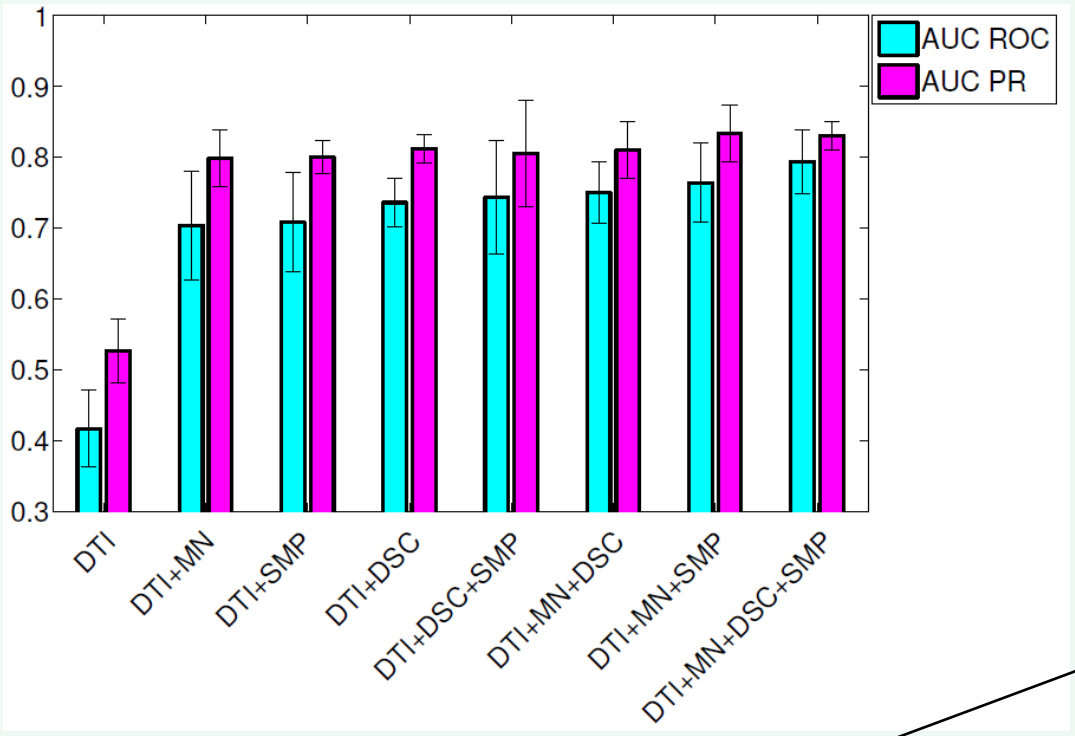
V. Gligorijevic, N. Malod-Dognin and **N. Przulj**, Patient-specific data fusion for cancer stratification and personalized treatment, *PSB*, 2016

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Patient-Specific Data Fusion → Personalized Treatment



Some results: 5-fold cross validation, average AUC: ROC and PR



Ovarian cancer patients:

1. Patient stratification → $\hat{\mathbf{C}}_1$
2. Driver gene prediction → $\hat{\mathbf{C}}_2$
3. **Drug repurposing** → $\hat{\mathbf{R}}_{23}$

$$\min_{\mathbf{G}_i \geq 0, 1 \leq i \leq 3} J = \min_{\mathbf{G}_i \geq 0, 1 \leq i \leq 3} \Big[\| \mathbf{R}_{12} - \mathbf{G}_1 \mathbf{H}_{12} \mathbf{G}_2^T \|_F^2 + \| \mathbf{R}_{23} - \mathbf{G}_2 \mathbf{H}_{23} \mathbf{G}_3^T \|_F^2 + tr(\mathbf{G}_2^T \mathbf{L}_2 \mathbf{G}_2) + tr(\mathbf{G}_3^T \mathbf{L}_3 \mathbf{G}_3) \Big]$$

V. Gligorijevic, N. Malod-Dognin and **N. Przulj**, Patient-specific data fusion for cancer stratification and personalized treatment, *PSB*, 2016

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Patient-Specific Data Fusion → Personalized Treatment

Some results: 37% of our ~225K predicted DTIs confirmed in MATADOR or CTD

Gene	Drug	Score	Clusters	DB
KIT	ATP	0.873	1, 2, 3	–
GABRQ	Adinazolam	0.808	1	M
GABRQ	Fludiazepam	0.808	1	M
GABRQ	Cinolazepam	0.809	1	M
GABRQ	Clotiazepam	0.809	1	M
HTR2A	Dopamine	0.809	1, 3	C, M
GRIN3A	Pethidine	0.801	1, 2	–
CACNA2D1	Verapamil	0.761	1, 3	M
PDGFRB	ATP	0.724	1, 2	–
KDR	ATP	0.724	1, 3	C
HTR1A	Mirtazapine	0.720	1, 2	C ,M
GABRA6	Adinazolam	0.688	1	M
GABRA6	Fludiazepam	0.688	1	M
GABRA6	Cinolazepam	0.688	1	M
GABRA6	Clotiazepam	0.688	1	M
GABRA4	Adinazolam	0.687	1, 3	M
GABRA4	Fludiazepam	0.687	1, 3	M
GABRA4	Cinolazepam	0.687	1, 3	M
GABRA4	Clotiazepam	0.687	1, 3	M
CACNA1D	Magnesium Sulfate	0.676	1, 2, 3	M

Ovarian cancer patients:

1. Patient stratification → $\hat{\mathbf{C}}_1$
2. Driver gene prediction → $\hat{\mathbf{C}}_2$
3. **Drug repurposing** → $\hat{\mathbf{R}}_{23}$

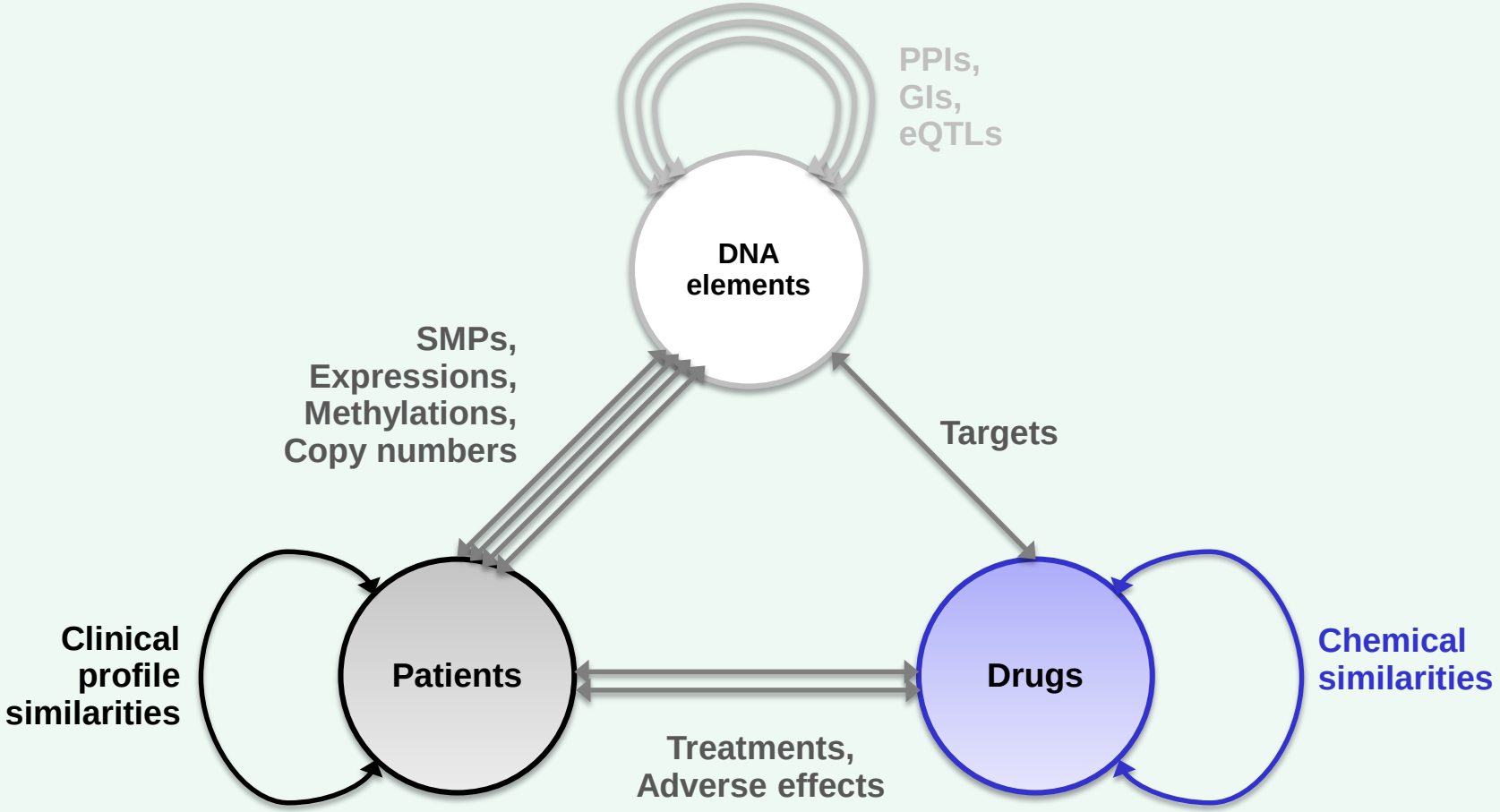
$$\min_{\mathbf{G}_i \geq 0, 1 \leq i \leq 3} J = \min_{\mathbf{G}_i \geq 0, 1 \leq i \leq 3} \Big[\| \mathbf{R}_{12} - \mathbf{G}_1 \mathbf{H}_{12} \mathbf{G}_2^T \|_F^2 + \| \mathbf{R}_{23} - \mathbf{G}_2 \mathbf{H}_{23} \mathbf{G}_3^T \|_F^2 + tr(\mathbf{G}_2^T \mathbf{L}_2 \mathbf{G}_2) + tr(\mathbf{G}_3^T \mathbf{L}_3 \mathbf{G}_3) \Big]$$

V. Gligorijevic, N. Malod-Dognin and **N. Przulj**, Patient-specific data fusion for cancer stratification and personalized treatment, *PSB*, 2016

2. Novel Methods

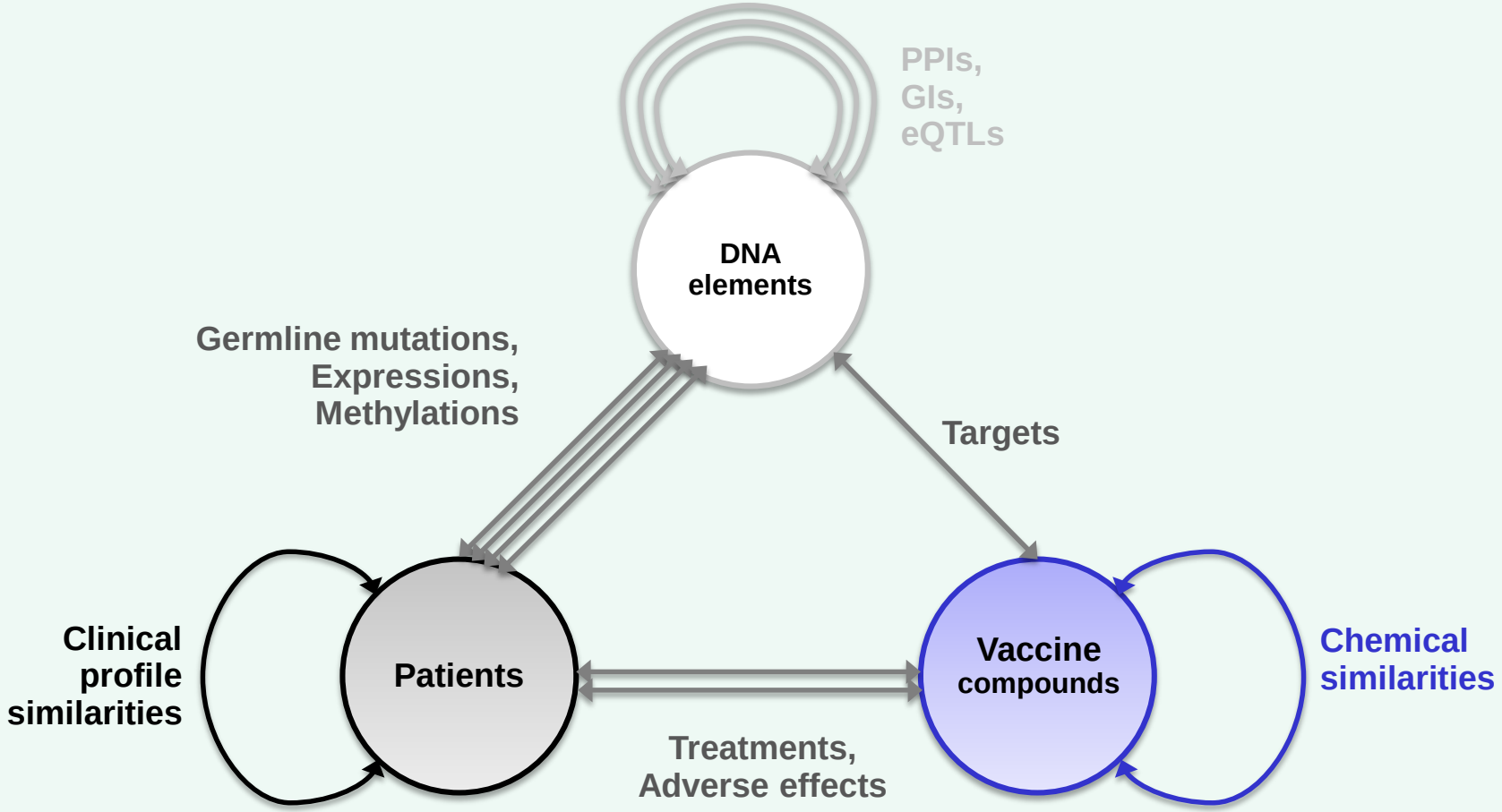
Mine the Medical World of Inter-Connected Entities

Patient-Specific Data Fusion → Personalized Treatment



2. Novel Methods
Mine the Medical World of Inter-Connected Entities

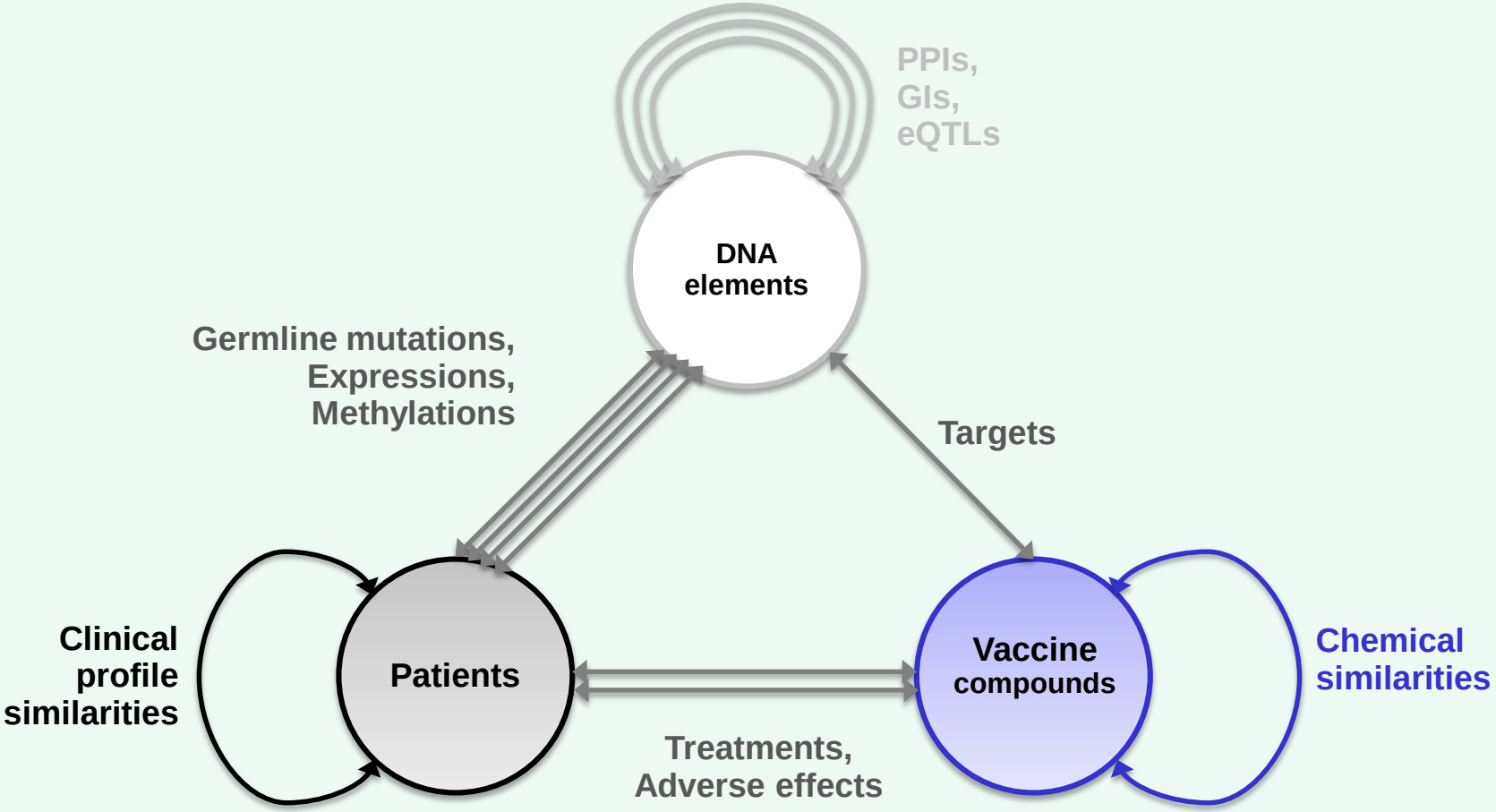
Patient-Specific Data Fusion → Personalized Treatment



- Systems vaccinology
  - With Dr. Nuria Izquierdo, IGTP IrsiCaixa, Badalona
  - Scientific Advisory Board of the Helmholtz Centre for Infection Research (HZI / Braunschweig, Germany)

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Patient-Specific Data Fusion → Personalized Treatment



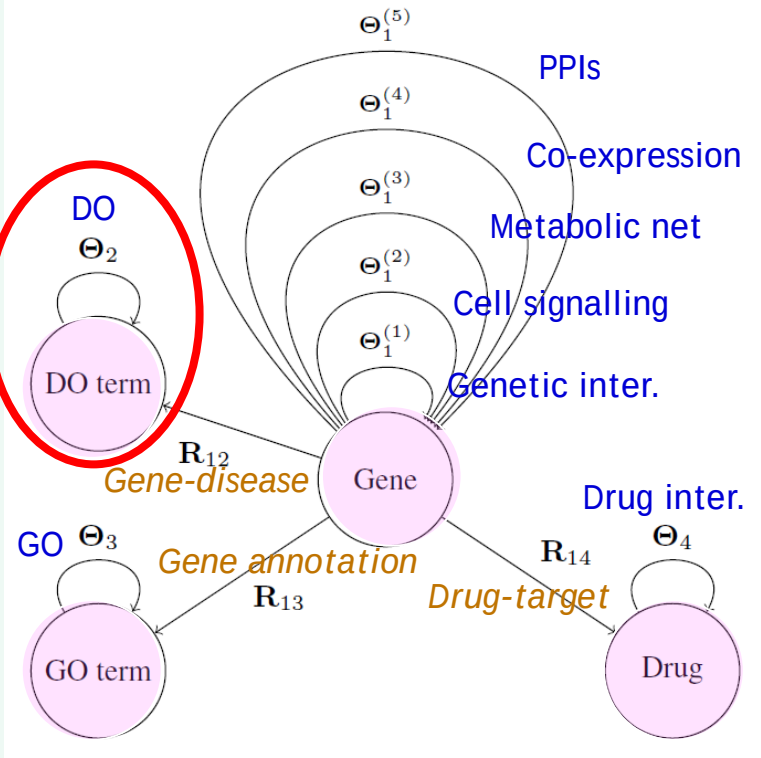
Obstacles:

1. **Different NP-hard continuous optimization problem:**
   - propose objective function,
   - optimization solver — prove convergence and correctness
2. **Optimization is slow → HPC**

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Disease Classification from Systems-Level Molecular Data

Method



4 Objects: Genes, GO terms, DO terms, Drugs
Constraints: Θ_i *(network topology, ontology relations)*
Relation matrices: R_{ij}

Some Results:

→ 14 disease-disease associations currently not present in DO:

- evidence for their relationships through *comorbidity* data and *literature curation*

→ GI the most important predictor of a link between diseases, despite small

→ Omission of any one of the included data sources reduces prediction quality

- Importance of systems-level data fusion

→ **DO ∩ disease class → 80% DO from *only* network data**

M. Zitnik, V. Janjic, C. Larminie , B. Zupan, and **N. Przulj,** Discovering disease-disease associations by fusing systems-level molecular data, *Scientific Reports - Nature*, 3:3202, 2013

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Disease Classification from Systems-Level Molecular Data

- Co-clustering GO terms, DO terms, Genes and Drugs under pairwise constraints:

$$\text{Input:}\quad \mathbf{R}=\begin{bmatrix} 0 & \mathbf{R_{12}} & \mathbf{R_{13}} & \mathbf{R_{14}} \\ \mathbf{R_{12}^T} & 0 & 0 & 0 \\ \mathbf{R_{13}^T} & 0 & 0 & 0 \\ \mathbf{R_{14}^T} & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{\Theta}=\begin{bmatrix} \mathbf{\Theta_1^{(t)}} & 0 & 0 & 0 \\ 0 & \mathbf{\Theta_2} & 0 & 0 \\ 0 & 0 & \mathbf{\Theta_3} & 0 \\ 0 & 0 & 0 & \mathbf{\Theta_4} \end{bmatrix}$$

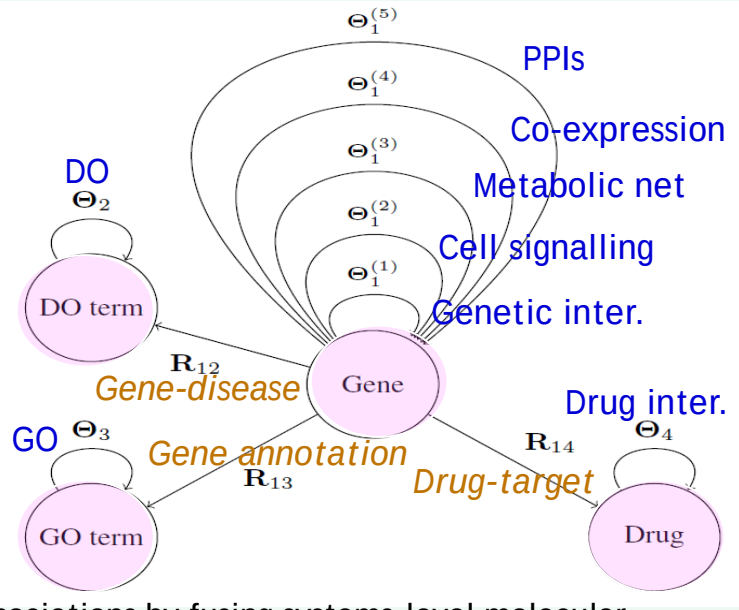
$$\text{Output:}\quad \mathbf{S}=\begin{bmatrix} 0 & \mathbf{S_{12}} & \mathbf{S_{13}} & \mathbf{S_{14}} \\ \mathbf{S_{21}} & 0 & 0 & 0 \\ \mathbf{S_{31}} & 0 & 0 & 0 \\ \mathbf{S_{41}} & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{G}=\begin{bmatrix} \mathbf{G_1} & 0 & 0 & 0 \\ 0 & \mathbf{G_2} & 0 & 0 \\ 0 & 0 & \mathbf{G_3} & 0 \\ 0 & 0 & 0 & \mathbf{G_4} \end{bmatrix}$$

- Minimizing Frobenious distance between R_{ij} and $G_i S_{ij} G_j^T$, for all relation matrices:
  - *i = {Genes}, j = {DO terms, GO terms, Drugs}*
  - G_i is a cluster indicator matrix for data type *i* (genes, DO terms, GO terms and Drugs)

with additional penalty terms:

$$\min_{\mathbf{G}\geq 0} J = \min_{\mathbf{G}\geq 0}\left[\parallel \mathbf{R}-\mathbf{GSG}^T \parallel_F^2 + \left(\sum_{t=1}^{5} tr(\mathbf{G}^T\mathbf{\Theta^{(t)}}\mathbf{G})\right)\right]$$

- Interested in G_2 (DO terms)
  - used for cluster assignment and inferring new disease associations from clusters

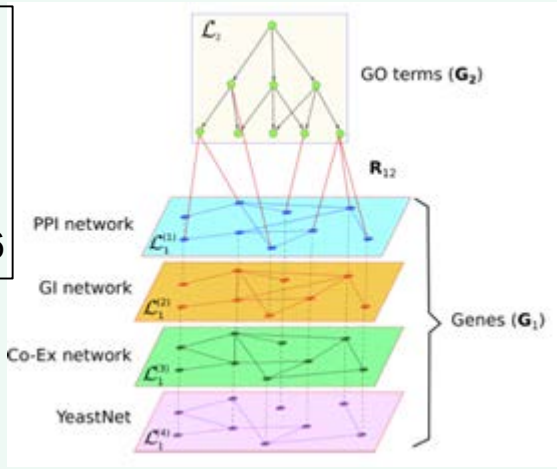


M. Zitnik, V. Janjic, C. Larminie , B. Zupan, and **N. Przulj,** Discovering disease-disease associations by fusing systems-level molecular data, *Scientific Reports - Nature*, 3:3202, 2013

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Gene Ontology from Systems-Level Molecular Data

- Outperform Dutkowski *et al*. [2013]
- 96% of GO reconstructed!
- Correct assignment of GO terms to genes (3-fold cross-validation, AUC=0.874±0.002)
- Graphlets improve results
- **Validated biologically** by Bonne's yeast Genetic Interaction profile data, *Science*, 2016



→ Optimization problem which minimizes $\| R_{12} - G_1 S_{12} G_2^T \|^2_F$ under the guidance of *pairwise constraints* (***connectivity*** *and* ***GDV similarity***) between genes in networks:

$$\min_{G_1 \geq 0, G_2 \geq 0} J = \min_{G_1 \geq 0, G_2 \geq 0} \left[\| R_{12} - G_1 S_{12} G_2^T \|_F^2 + \left(\sum_{i=1}^{4} tr(G_1^T L_1^{(i)} G_1) \right) + \left(\sum_{i=1}^{4} tr(G_1^T \Lambda_1^{(i)} G_1) \right) + tr(G_2^T L_2 G_2) \right]$$

→ using topology of molecular networks as constraints (penalty terms) in this optimization problem:

→ $L_1^{(i)}$ is Laplacian of **adjacency matrix** of a molecular network *i=1,2,3,4:*
$L_1^{(i)} = D^i - A^i$, D^i is diagonal matrix of degrees (row summation of A^i), A^i is adjacency matrix

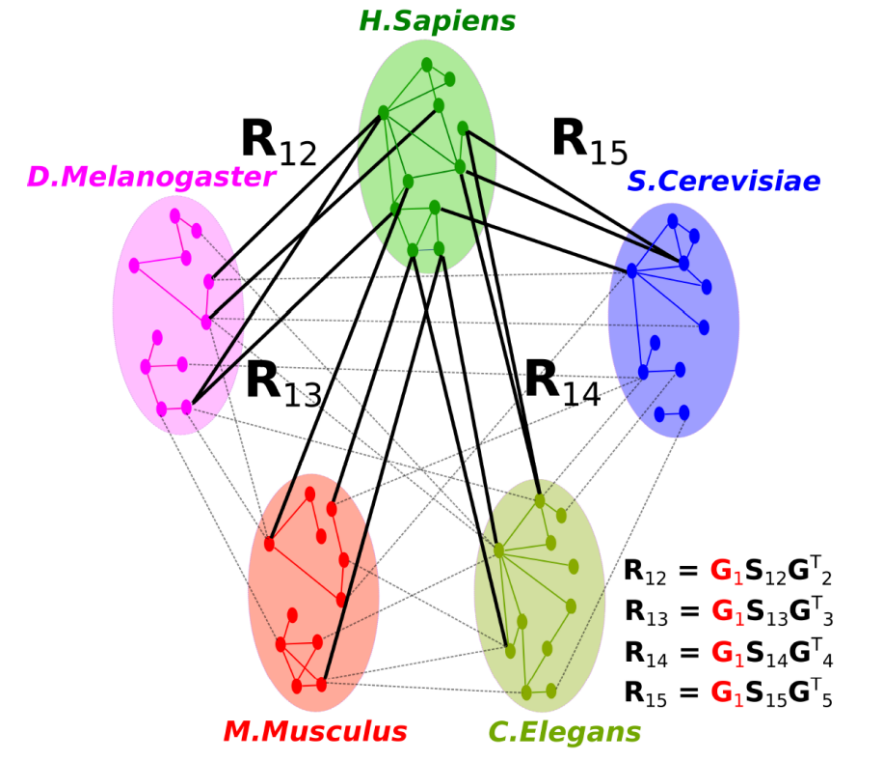
→ $\Lambda_1^{(i)}$ are Laplacians of **GDV similarity matrices** over all genes for each molecular network *i*:
$\Lambda_1^{(i)} = D^i - \sigma^{(i)}$, D^i is diagonal matrix of row summation of $\sigma^{(i)}$, $\sigma^{(i)}$ is binary GDV similarity matrix (containing only significantly similar gene/protein pairs)

→ L_2 is Laplacian of **Gene Ontology graph**

V. Gligorijevic, V. Janjic and **N. Przulj**, Integration of molecular network data reconstructs GO, *Bioinformatics, Vol. 30 ECCB 2014, i594–i600 (14% acceptance rate),* 2014

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Multiple Network Alignment: ***Fuse***



Algorithm 1 Approximate maximum weight k-partite matching.

Input $G = (\bigcup_{i=1}^{k} V_i, E, W)$
for $i = \{2, \ldots, k\}$ **do**
  Find maximum weight bipartite matching $F_{1,i}$ of $G[V_1, V_i]$
  Construct G_{1i}, the merge of V_1 and V_i from G along $F_{1,i}$
  Set $G = G_{1i}$, and relabel V_{1i} as V_1
$C = \{\emptyset\}$
for each merged node u in V_1 **do**
  Cluster C_u is the set of nodes that are merged into u
  Add C_u to C
Output C

We use a block-based representation of relation ($\mathbf{R}$) and Laplacian ($\mathbf{L}$) matrices and matrix factors ($\mathbf{S}$ and $\mathbf{G}$) for our 5 PPI networks as follows:

$$\mathbf{R} = \begin{bmatrix} 0 & \mathbf{R}_{12} & \ldots & \mathbf{R}_{15} \\ \mathbf{R}_{12}^T & 0 & \ldots & \mathbf{R}_{25} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{15}^T & \mathbf{R}_{25}^T & \ldots & 0 \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} \mathbf{L}_1 & 0 & \ldots & 0 \\ 0 & \mathbf{L}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{L}_5 \end{bmatrix};$$

$$\mathbf{S} = \begin{bmatrix} 0 & \mathbf{S}_{12} & \ldots & \mathbf{S}_{15} \\ \mathbf{S}_{12}^T & 0 & \ldots & \mathbf{S}_{25} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{15}^T & \mathbf{S}_{25}^T & \ldots & 0 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & 0 & \ldots & 0 \\ 0 & \mathbf{G}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{G}_5 \end{bmatrix}$$

To simultaneously factorize all relation matrices, $\mathbf{R}_{ij} \approx \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T$, $0 \leq i, j \leq 5$, under the constraints of PPI networks, we minimize the following objective function:

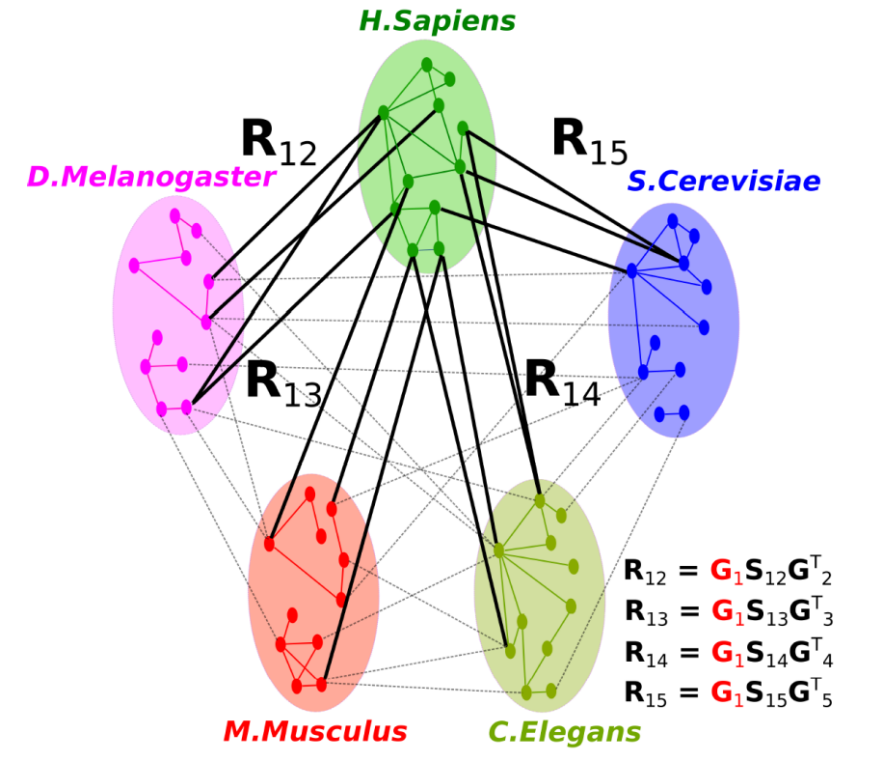
$$\min_{\mathbf{G} \geq 0} J = \left[\| \mathbf{R} - \mathbf{G}\mathbf{S}\mathbf{G}^T \|_F^2 + \gamma Tr(\mathbf{G}^T \mathbf{L} \mathbf{G}) \right] \quad (2)$$

where Tr denotes the trace of a matrix and γ is a regularization parameter which balances the influence of network topologies in reconstruction of the relation matrix. The second term of equation 2 is the penalization term.

V. Gligorijevic, N. Malod-Dognin and **N. Przulj,** Fuse: Multiple network alignment via data fusion, *Bioinformatics*, 32(8):1195-203, 2016. IF=7.3

2. Novel Methods
Mine the Medical World of Inter-Connected Entities

Multiple Network Alignment: ***Fuse***



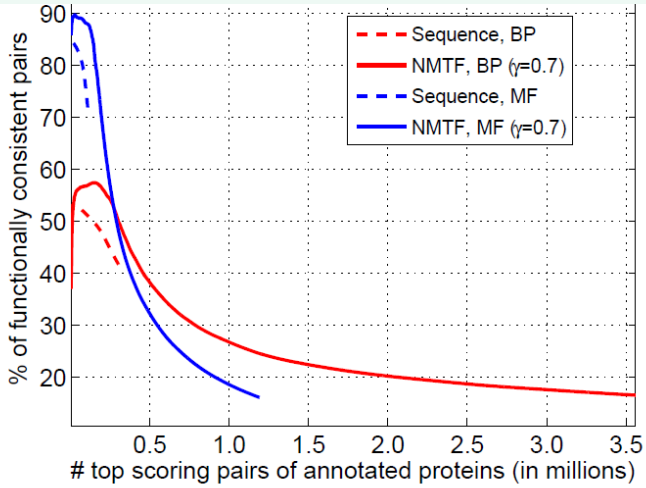


Fig. 2. Functional consistency of NMTF associations. For both NMTF associations and sequence similarity of protein pairs, we plot the cumulative number of protein pairs with both proteins annotated (x-axis) against the percentages of them sharing GO terms (y-axis). Biological process (BP) and molecular function (MF) annotations are considered separately.

V. Gligorijevic, N. Malod-Dognin and **N. Przulj,** Fuse: Multiple network alignment via data fusion, *Bioinformatics*, 32(8):1195-203, 2016. IF=7.3

Overview

Medicine: complex world of inter-connected entities

1. **Motivation**
2. **New Methods – Examples:** mine inter-connected data
   i. **Single layer of omics data: Molecular networks** → function, disease
   ii. **Multiple layers of heterogeneous data:**
      - **Patient-centered data integration → Precision medicine**
      - **Disease re-classification**
      - **Gene Ontology reconstruction**
      - **Network alignment**
3. **Vision**

3. Vision

<u>Biomedical Data</u>: complex system of heterogeneous interacting entities

- Large
- Heterogeneous
- Highly dimensional
- Growing Complexity
- Noisy
- Dynamic
- Different time and space scales

- World Economic Forum in Davos 2016:
  - "Big data" potential to transform medicine
  - Make it more effective due to increased life expectancy and exposure to environmental risks
- *Nature* Insight and Outlook of 2015 and 2016

- I was awarded 2014 BCS Roger Needham Award in recognition of "the potential my work and research have to revolutionize health and pharmaceutics"

3. Vision

<u>Biomedical Data</u>: complex system of heterogeneous interacting entities

- Large
- Heterogeneous
- Highly dimensional
- Growing Complexity
- Noisy
- Dynamic
- Different time and space scales

- Each type: ***limited***, but ***complementary*** information
- **Seek principled, joint organization and mining within the same framework**

- World Economic Forum in Davos 2016:
  - "Big data" potential to transform medicine
  - Make it more effective due to increased life expectancy and exposure to environmental risks
- *Nature* Insight and Outlook of 2015 and 2016

- I was awarded 2014 BCS Roger Needham Award in recognition of "the potential my work and research have to revolutionize health and pharmaceutics"

€2M **ERC Consolidator Grant** for 2018-2023
Title: "Integrated Connectedness for a New Representation of Biology"

3. Vision

Holistically Mine All Available Data

→ **Paradigm shifts**

1. **Conceptual**
2. **Methodological**

N. Przulj and N. Malod-Dognin, “Network analytics in the age of Big Data,” ***Science*** 353:6295, 2016

3. Vision

Holistically Mine All Available Data

→ **Paradigm shifts**

1. Conceptual

Do not analyze single data type in isolation of others (e.g., sequence align.)

- **Analyze all types of data within a single framework**
- **New, bottom-up, data-driven biological concepts**
  - Elucidate that a cell may be governed by yet undiscovered principles of life
  - Point to ways to re-think biology and approaches to medicine

3. Vision

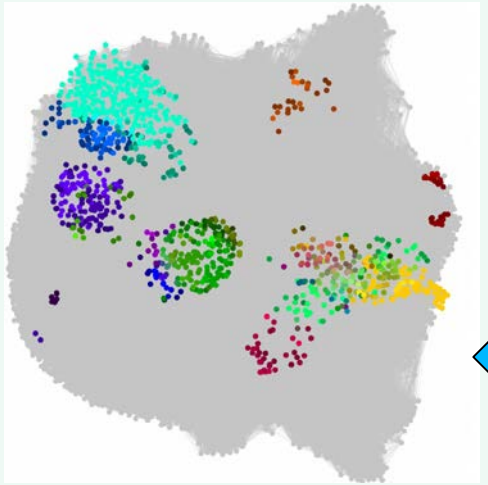
Holistically Mine All Available Data

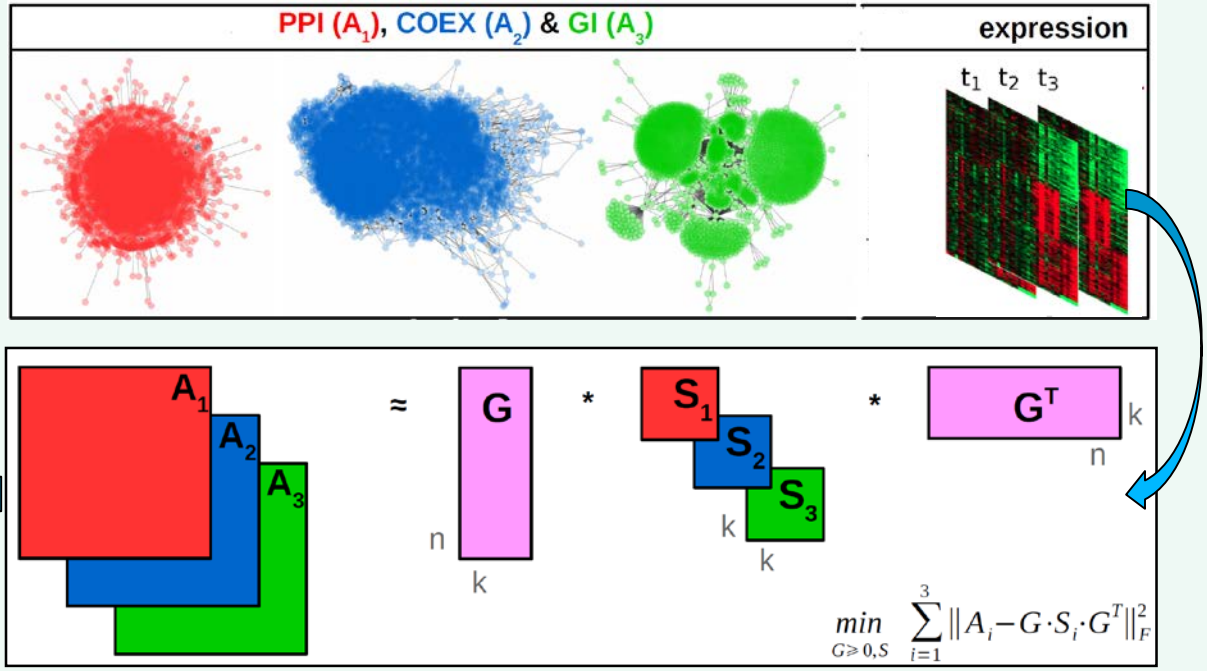
→ **Paradigm shifts**

1. Conceptual

Do not analyze single data type in isolation of others (e.g., sequence align.)

- **Introduce a concept of an "Integrated Cell (*iCell*)"**





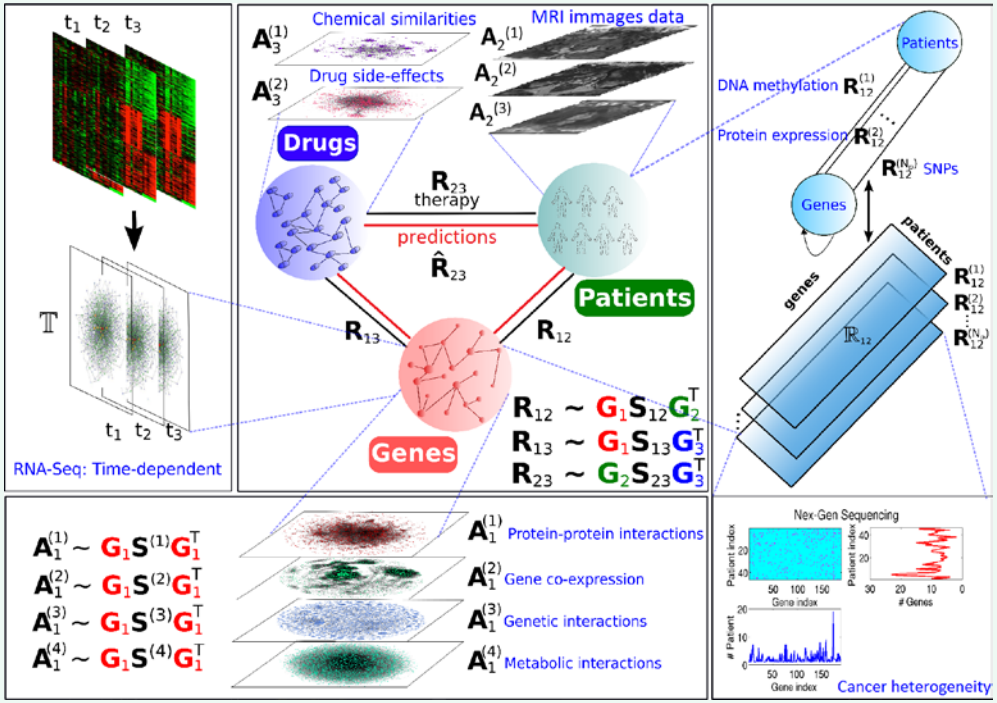
Noël Malod-Dognin, Julia Petschnigg, Sam F. L. Windels, Janez Povh, Harry Hemmingway, Robin Ketteler and **Nataša Pržulj**, "iCell: integrated cells uncover new cancer genes," ***Nature Communications***, 2019

3. Vision

Holistically Mine All Available Data

→ **Paradigm shifts**

2. **Methodological**



N. Przulj and N. Malod-Dognin, "Network analytics in the age of Big Data," ***Science*** 353:6295, 2016

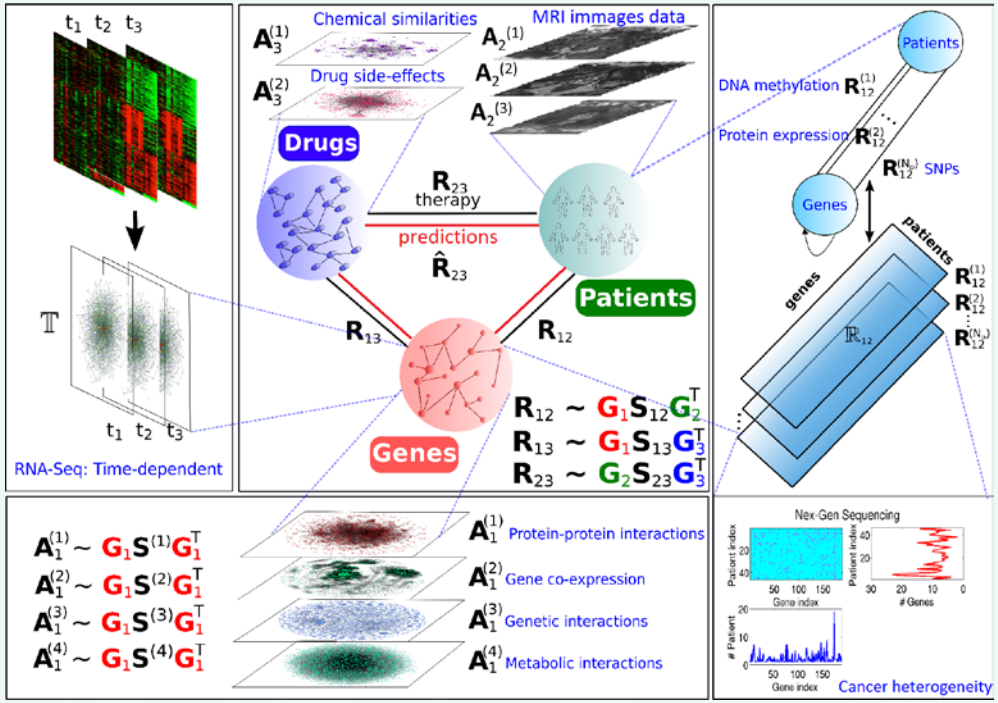
3. Vision

Holistically Mine All Available Data

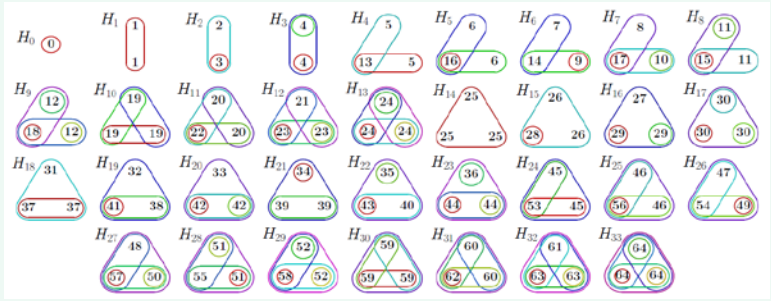
→ **Paradigm shifts**

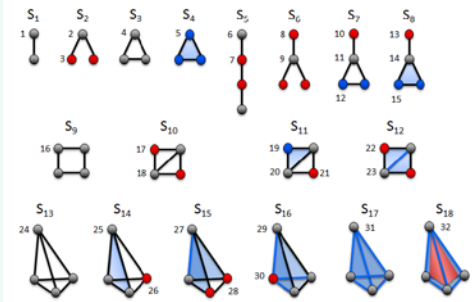
2. Methodological

- **Mathematical formalisms**
  - Capture **multi-scale organization**
  - **Dynamics, stochasticity** of the data,…
  - E.g., multiplex networks, hypergraphs, simplicial complexes …



- **Algorithms** to compute and **extract information** from those formalisms





T. Gaudelet, N. Malod-Dognin and **N. Przulj**, "Higher order molecular organisation as a source of biological function," *Bioinformatics*, ECCB'18

N. Malod-Dognin and **N. Przulj**, "Functional geometry of protein-protein interaction networks," arXiv:1804.04428, 2018

Noël Malod-Dognin, Julia Petschnigg, Sam F. L. Windels, Janez Povh, Harry Hemmingway, Robin Ketteler and **Nataša Pržulj**, "iCell: integrated cells uncover new cancer genes," ***Nature Communications***, 2019

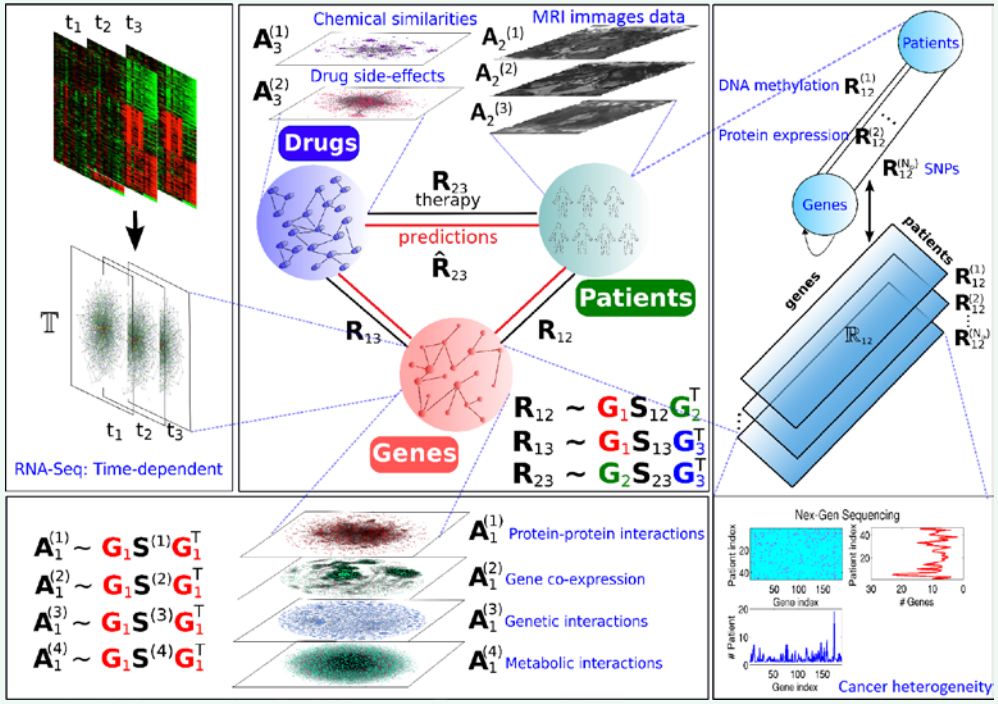
3. Vision

Holistically Mine All Available Data

→ **Paradigm shifts**

2. Methodological

- **Mathematical formalisms**
  - Capture **multi-scale organization**
  - **Dynamics, stochasticity** of the data,…
  - E.g., multiplex networks, hypergraphs, simplicial complexes …



- **Algorithms** to compute and **extract information** from those formalisms

How: e.g.

- Utilize dependencies in local network topology (orbits) — data set dependent
- Uncover latent low-dimensional structure of data
- Exploit structure for developing efficient toolsets for particular data

N. Przulj and N. Malod-Dognin, "Network analytics in the age of Big Data," ***Science*** 353:6295, 2016

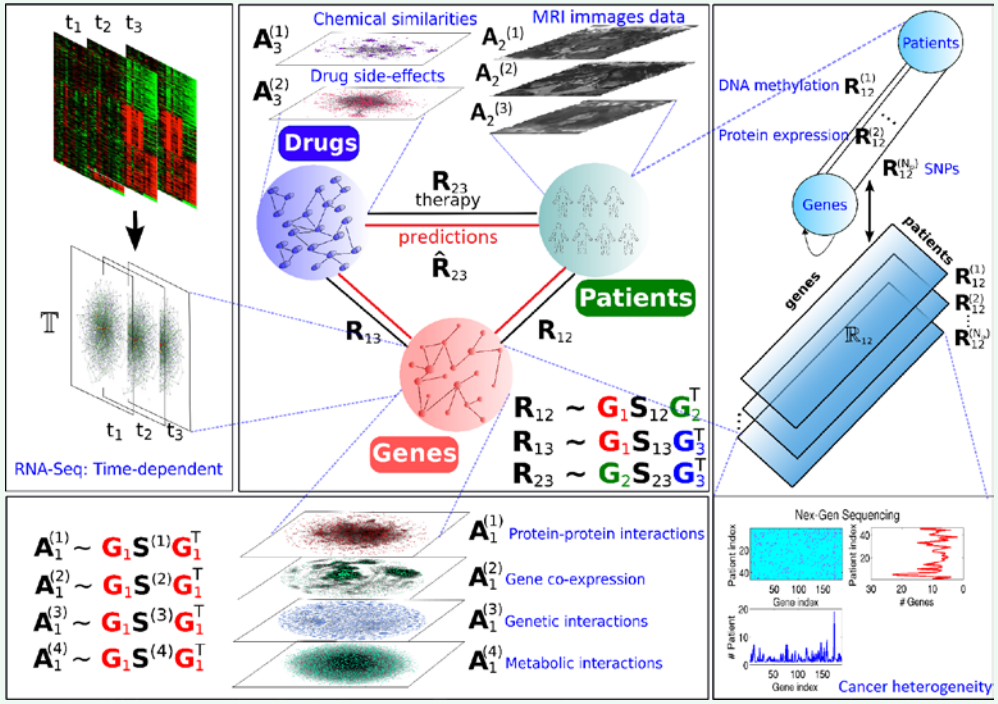
3. Vision

Holistically Mine All Available Data

→ **Paradigm shifts**

2. Methodological

- **Mathematical formalisms**
  - Capture **multi-scale organization**
  - **Dynamics, stochasticity** of the data,…
  - E.g., multiplex networks, hypergraphs, simplicial complexes …



- **Algorithms** to compute and **extract information** from those formalisms

Computational issues remain to be addressed, arising from intractability:

- large sizes, complexity, heterogeneity, noisiness, and
- different time and space scales of the data

"Embedded" data scientists: problem-specific heuristic methods, HPC

N. Przulj and N. Malod-Dognin, "Network analytics in the age of Big Data," ***Science*** 353:6295, 2016

3. Vision

Holistically Mine All Available Data

→ **Paradigm shifts**

Guided by Needs of Biomedical Collaborators and Industry

E.g.:

- Cancer
- Rare genetic diseases
- Viral medicines
- JnJ
- GSK
- Medium, start-ups, …

Acknowledgements

- **Funding:**

European Research Council
Funding for top researchers from anywhere in the world

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

The Farr Institute of Health Informatics Research

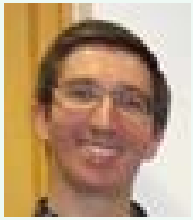
the PROSTATE project

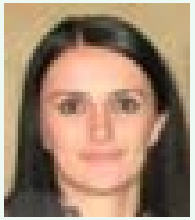
SLOVENIAN RESEARCH AGENCY

Република Србија
МИНИСТАРСТВО ПРОСВЕТЕ НАУКЕ И ТЕХНОЛОШКОГ РАЗВОЈА

- **Group members (present and past):**

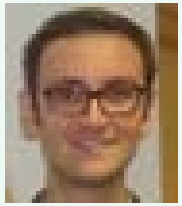
1. Dr. Noel Malod-Dogning
2. Dr. Julia Petschnigg
3. Dr. Chhedi Gupta
4. Sam Windels
5. Thomas Gaudelet
6. Dr. Omer Yaveroglu
7. Prof. Tijana Milenković
8. Dr. Oleksii Kuchaiev
9. Dr. Vesna Memišević
10. Dr. Vladimir Gligorijevic












- **Collaborators:**

Robin Ketteler, Harry Hemmingway, Igor Stagljar, Charles Boone, …

The Chartered Institute for IT

ACADEMIA EUROPAEA } The Academy of Europe

Acknowledgements

➢ **Funding:**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

European Research Council
Funding for top researchers from anywhere in the world

NSF

NSERC CRSNG

Ontario

gsk

IBM

Google

J&J

The Farr Institute of Health Informatics Research

the PROSTATE project

arrs
SLOVENIAN RESEARCH AGENCY

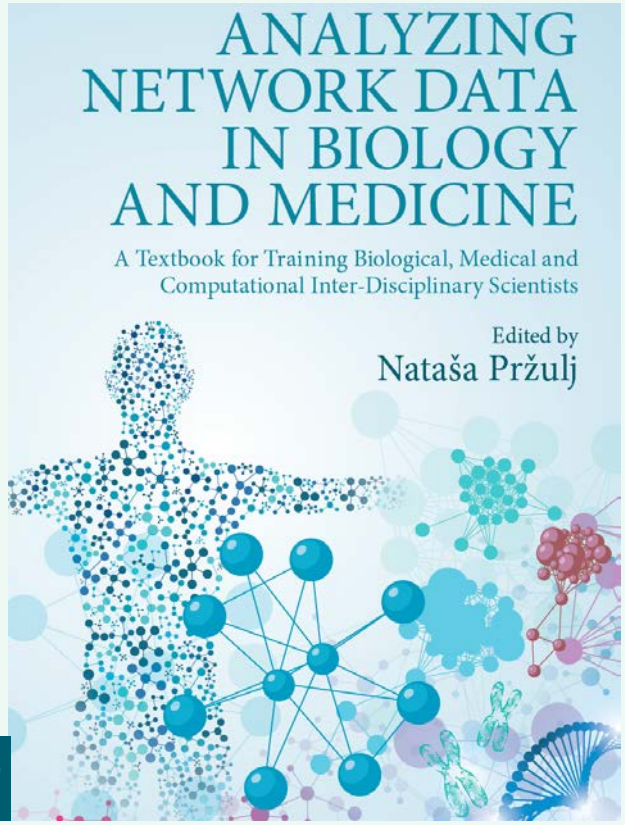
Република Србија
МИНИСТАРСТВО ПРОСВЕТЕ НАУКЕ И ТЕХНОЛОШКОГ РАЗВОЈА

ERC Consolidator Grant:

- Post-Doc positions
- PhD student positions

JnJ:

- Post-Doc position



Thank you

Comments and Questions