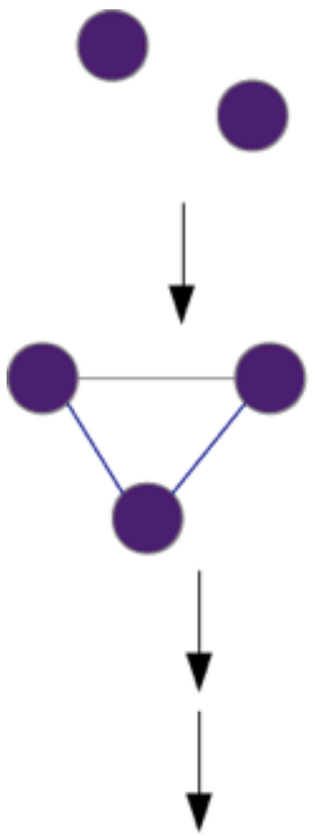
Marseille Medical Genetics

Mining Networks to study Rare Common Diseases

Anaïs Baudot
anais.baudot@univ-amu.fr

BSC, 31/01/2018

Genotype

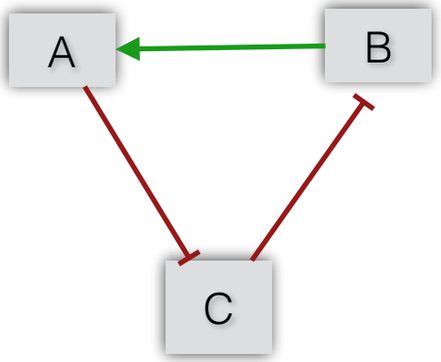


Networks !!

Phenotype

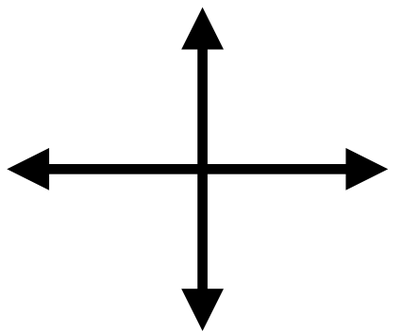
Complex Systems
Systems Biology

NSBD



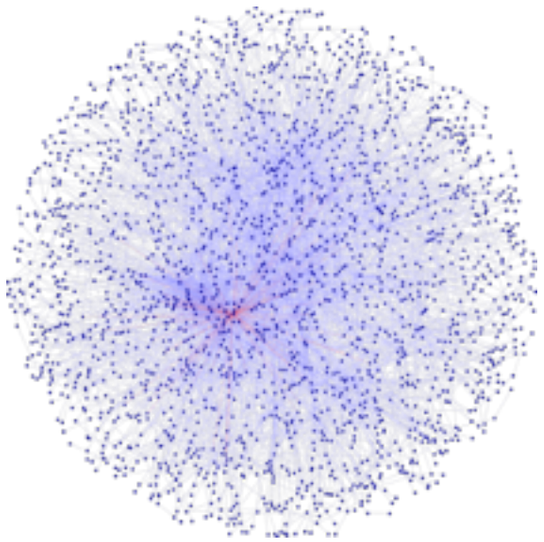
Dynamical Network Modeling

Algorithmic developments

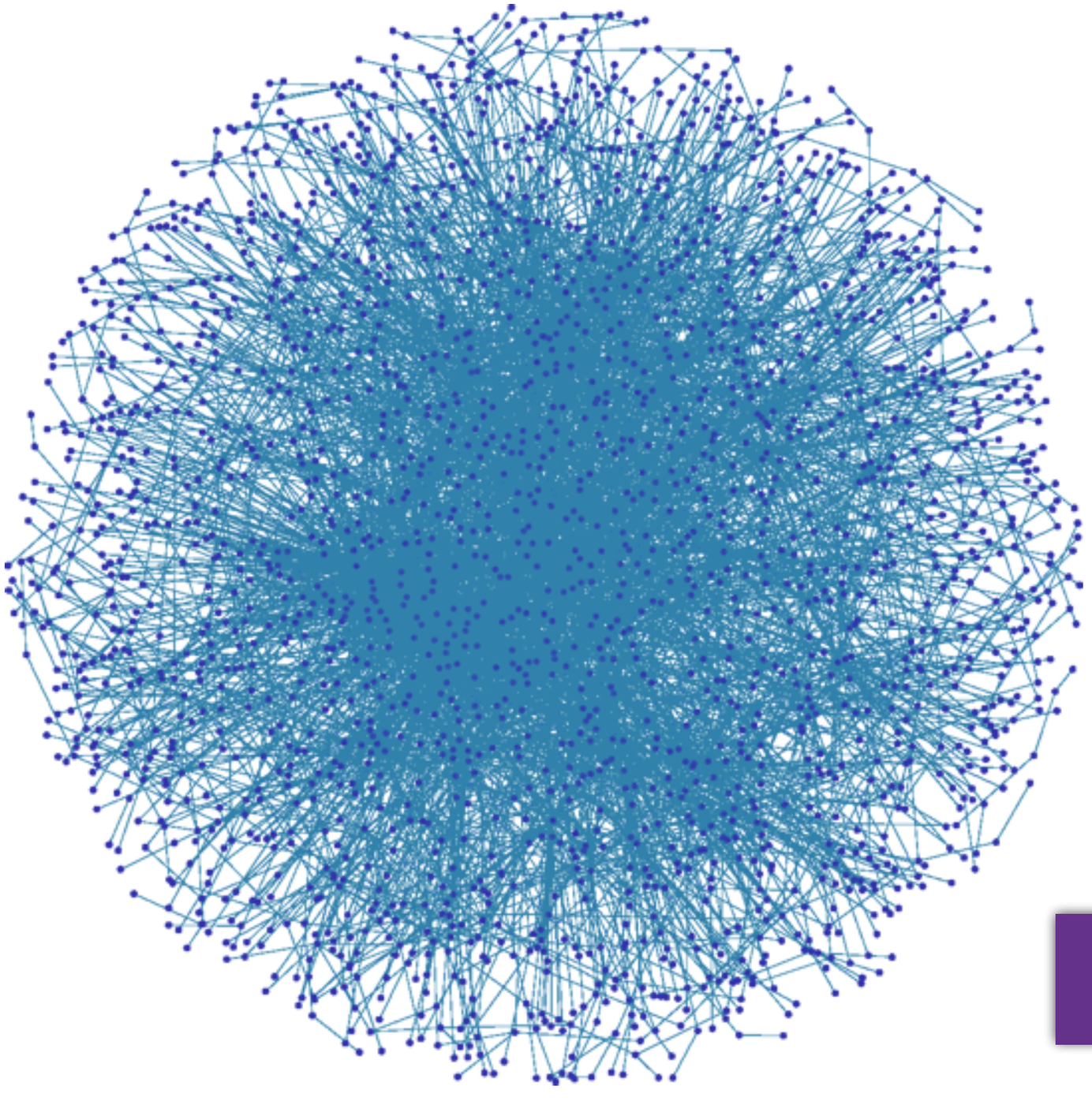


Applications to human diseases

Large-scale Network Mining

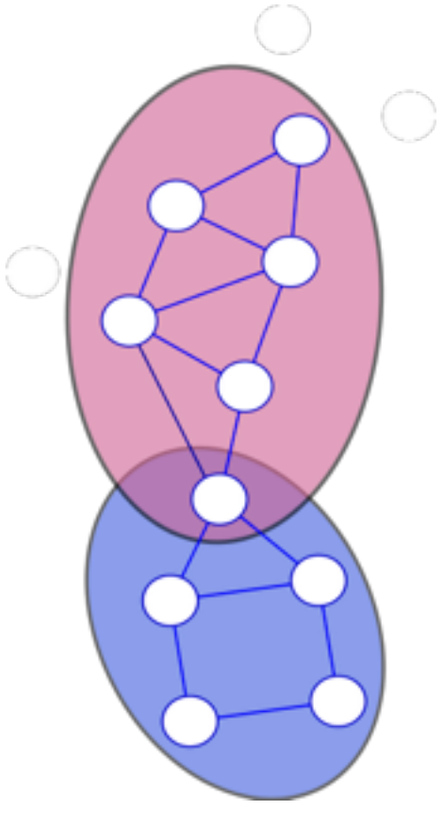


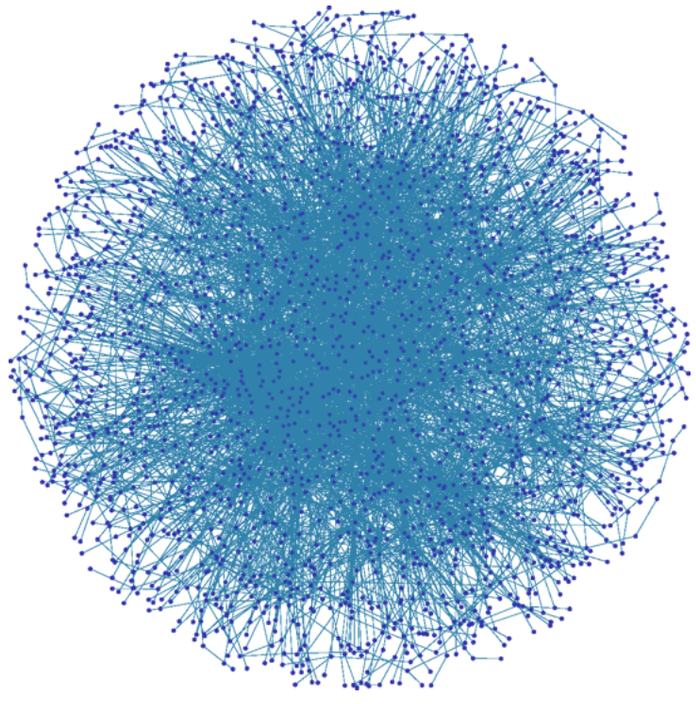
Interactome: The set of <u>known & possible</u> protein-protein interactions



~ 60 000 binary undirected interactions
~ 12 000 human proteins

How to extract relevant biological information ?



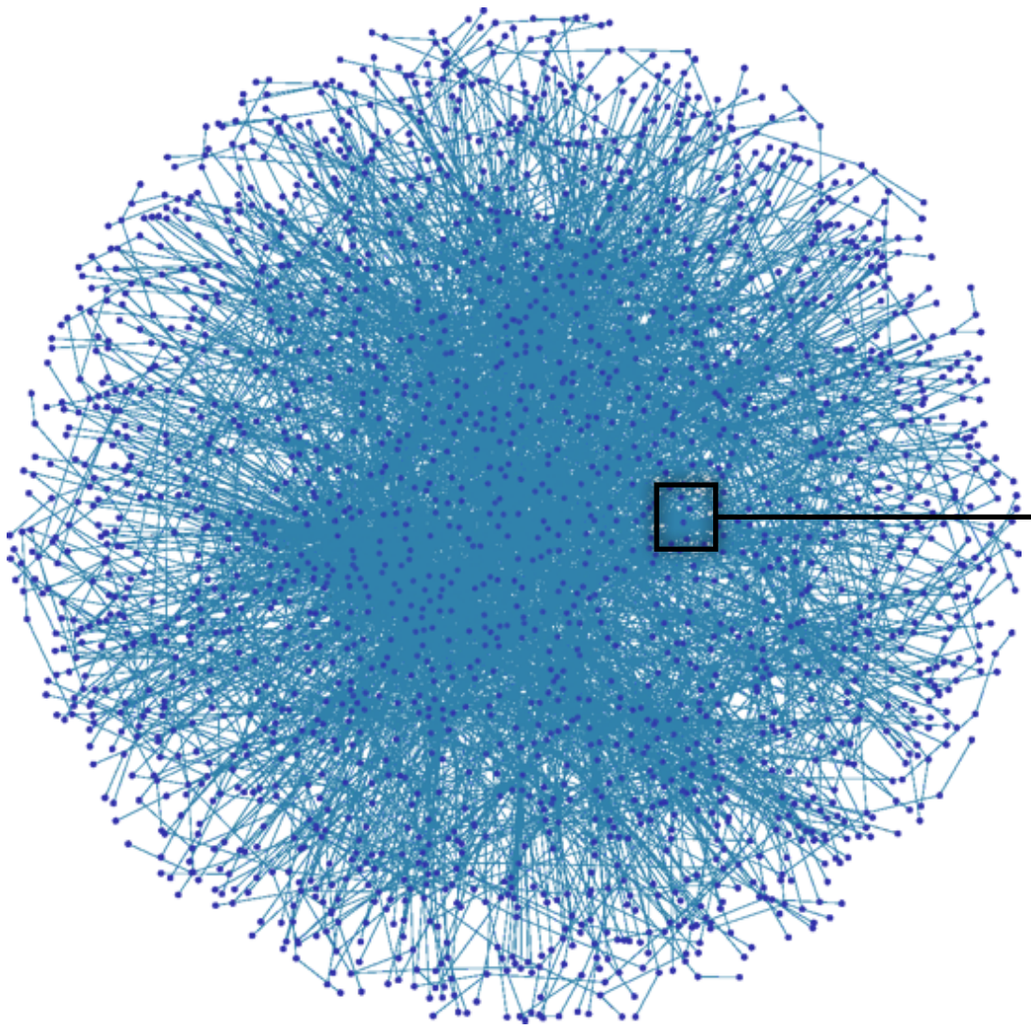


Binary interaction **Functional modules** **Interaction Networks**

Precision **Comprehensiveness**

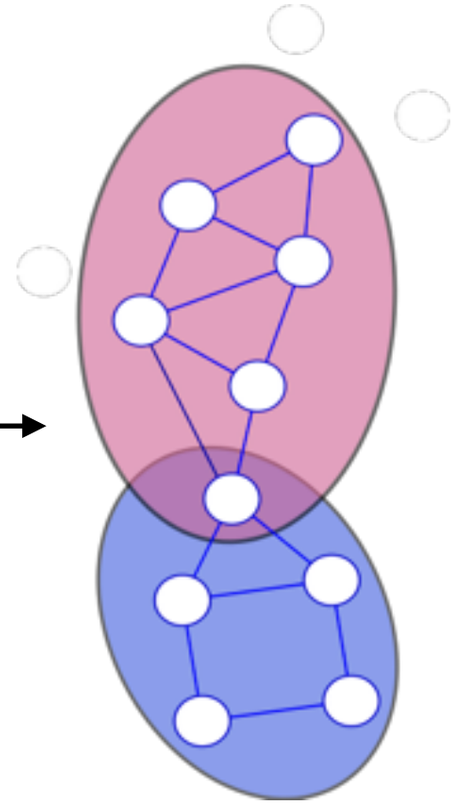
Inspired from P. Aloy, ECCB 2014

Community identification algorithms



Topology : Density,
Modularity optimization

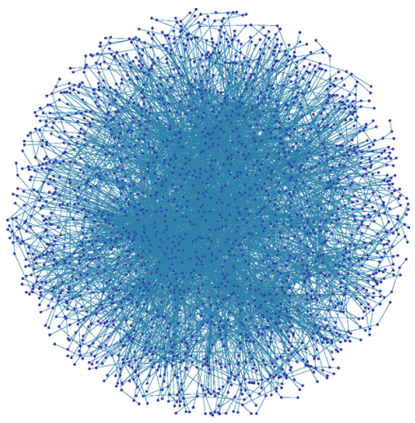
Overlapping / Disjoint



Spinelli et al. (2013)

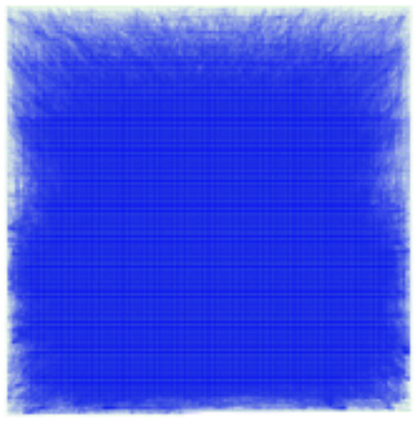
Many Biological Networks

PPI



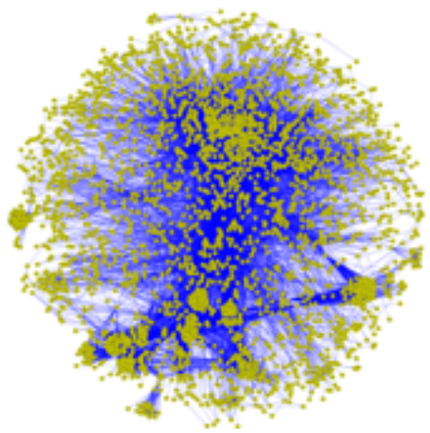
~60 000 edges

Co-expr



~1 400 000 edges

Pathways



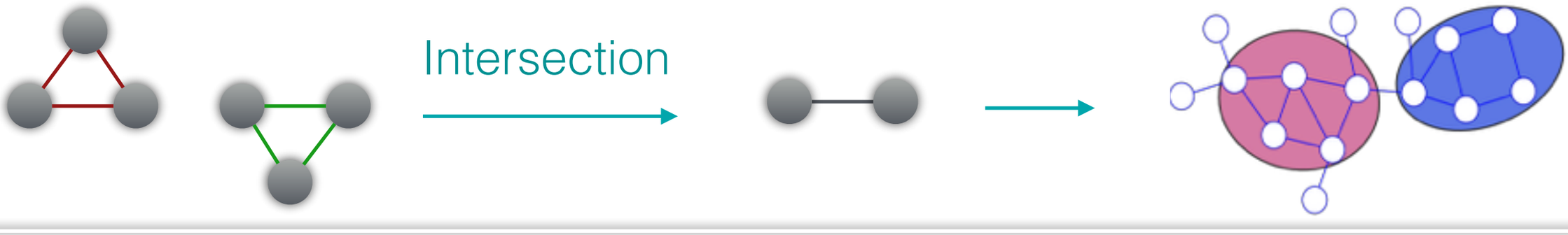
~250 000 edges

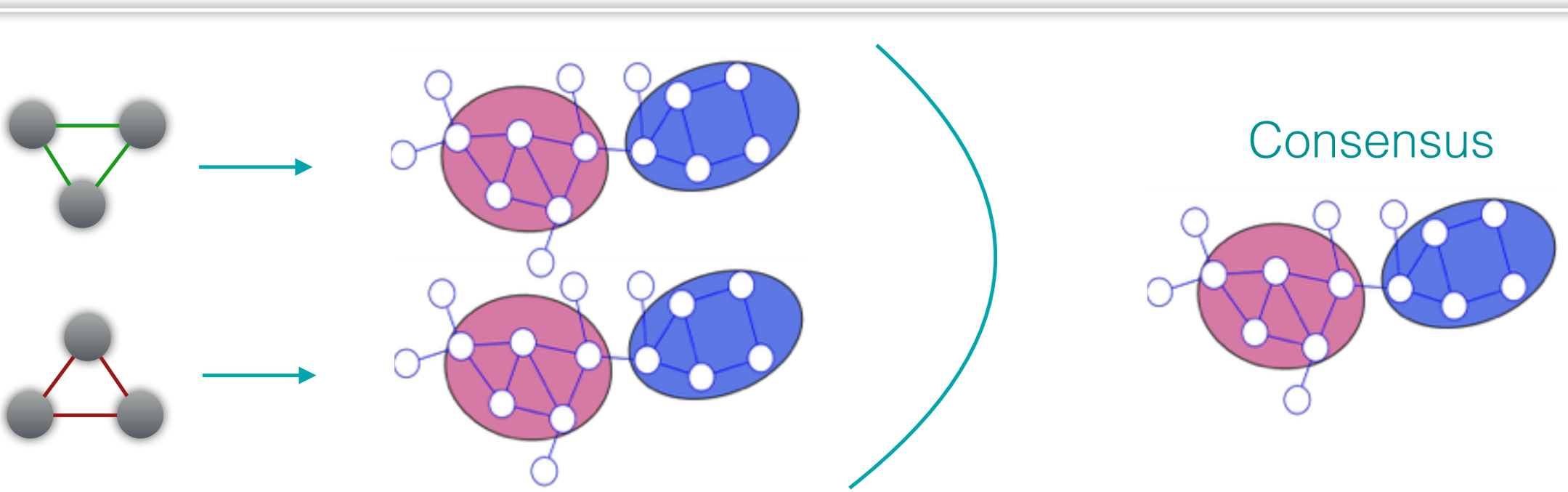
Complexes

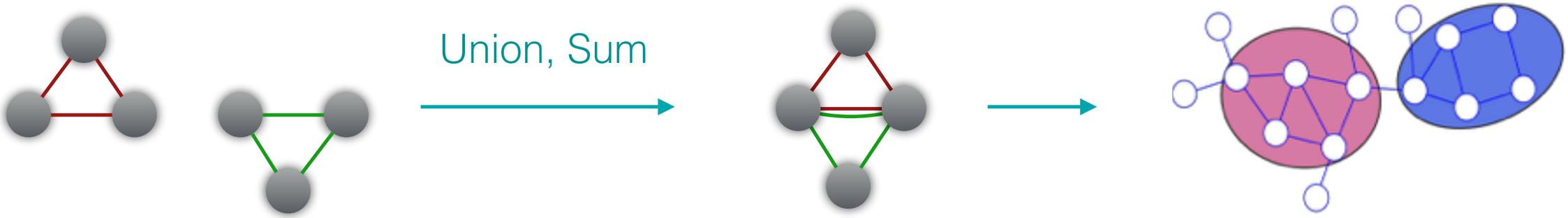
~40 000 edges

How do we combine many network/ interaction sources?

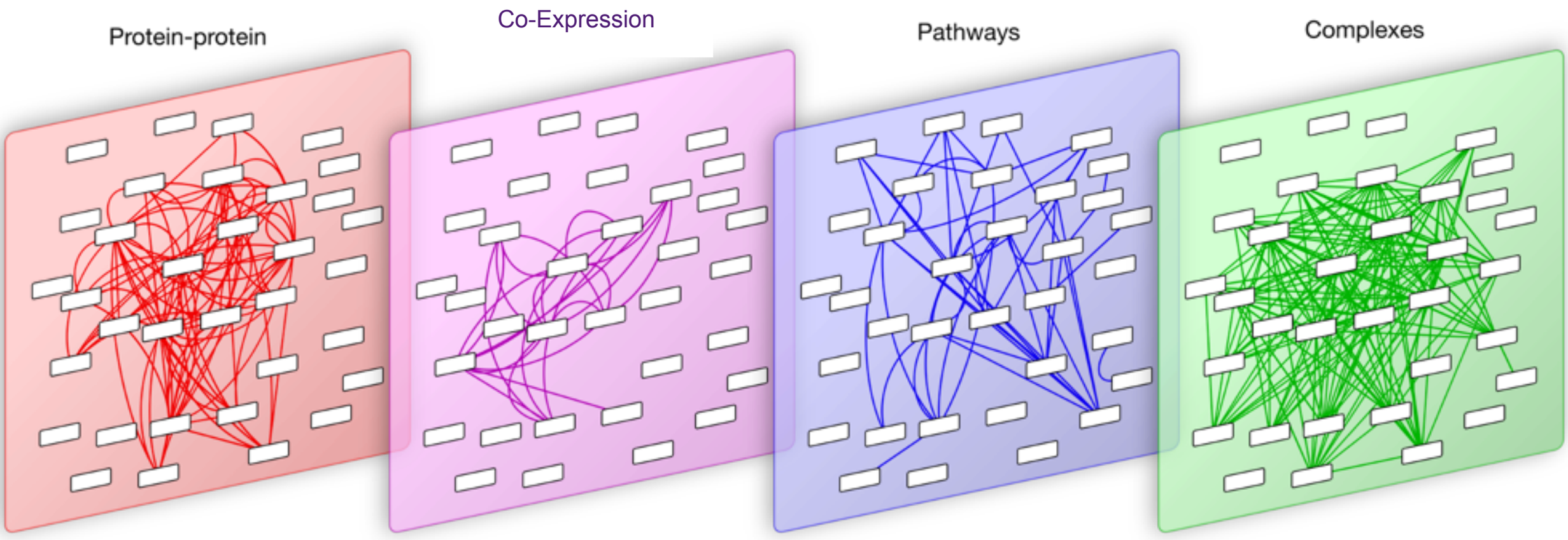
Communities from many Networks





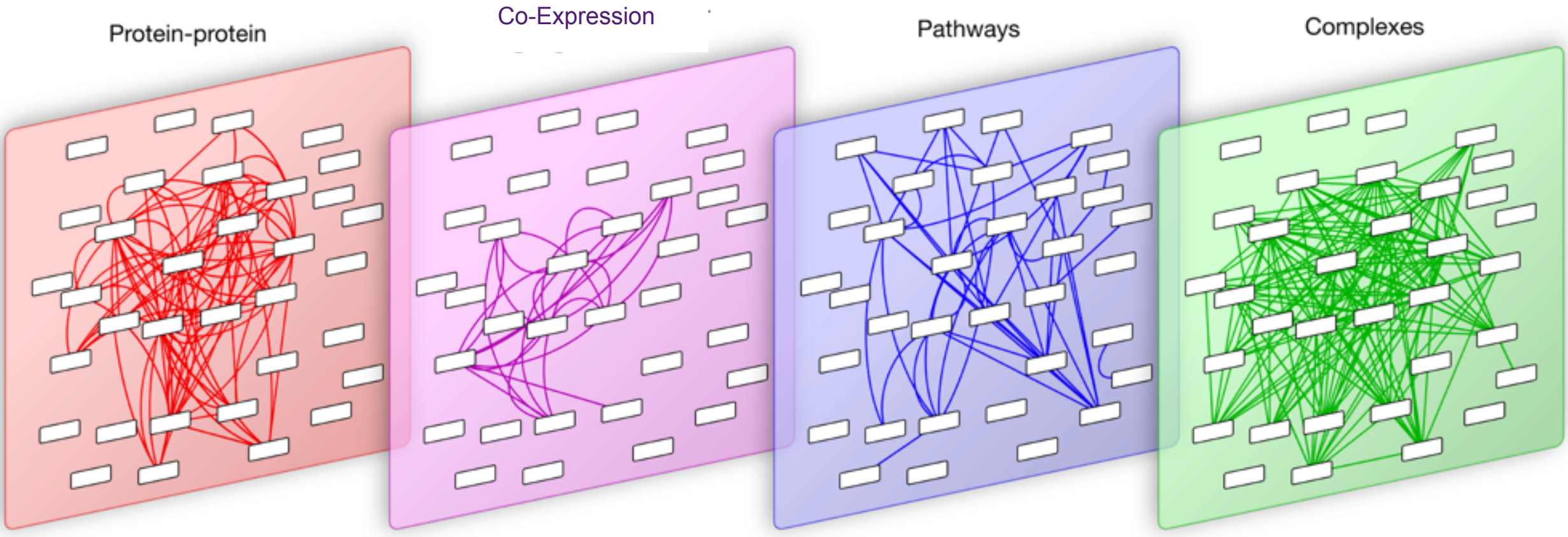


Multiplex framework

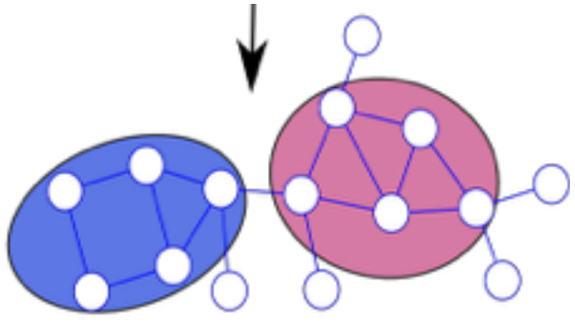


Multiplex, Multi-layer, Multi-slice

Multiplex framework

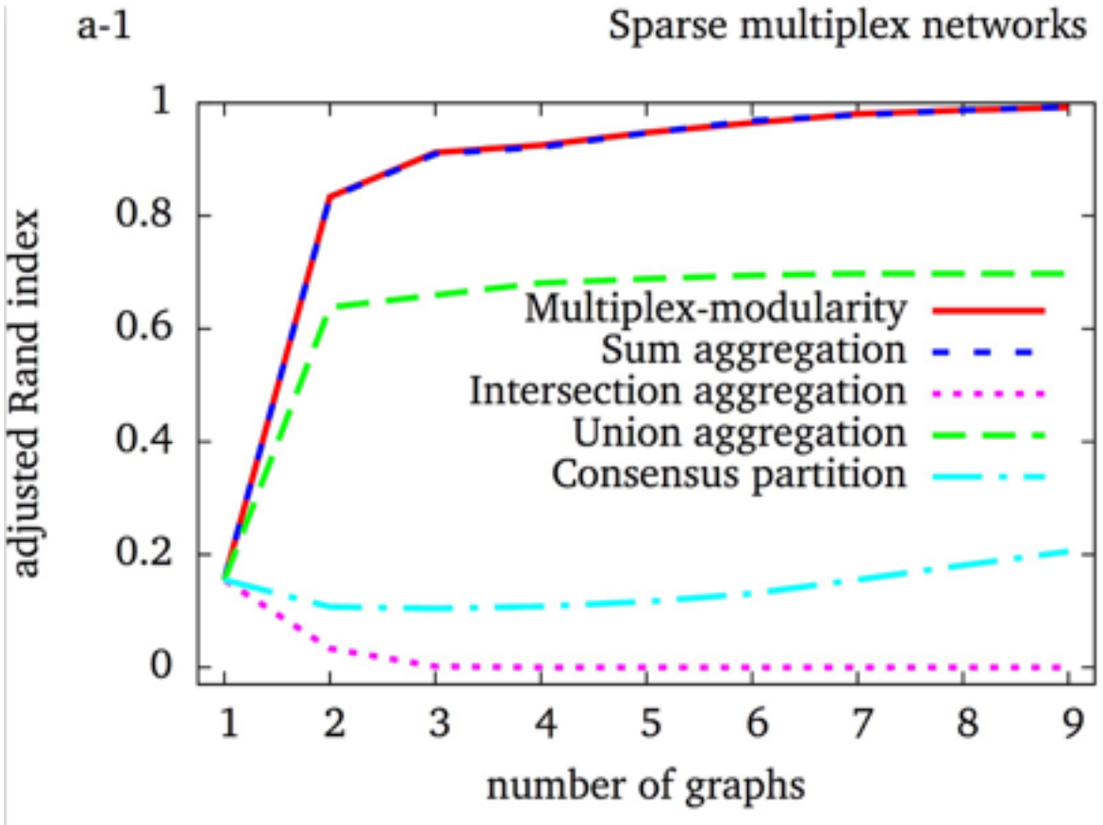


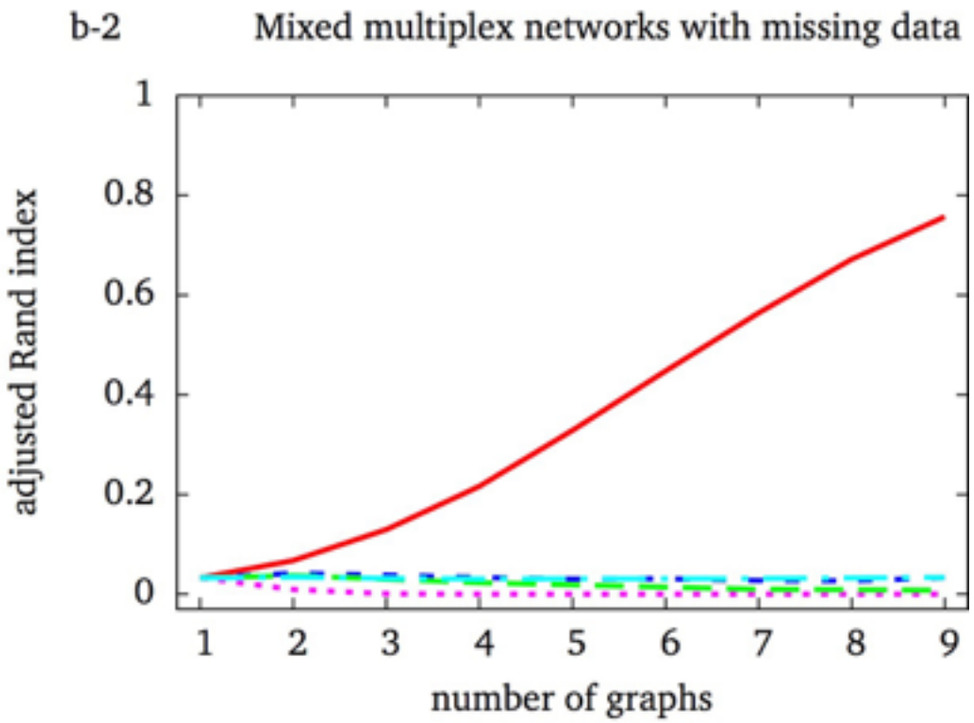
Multiplex-modularity



$$Q^M((X^{(g)})_g, \mathbf{c}) = \sum_a \left[\sum_g \frac{\sum_{\{i,j\}} X_{i,j}^{(g)}}{2m^g} - \sum_g \frac{\sum_{\{i,j\}} k_i^g k_j^g}{(2m^g)^2} \right]$$

Simulated networks

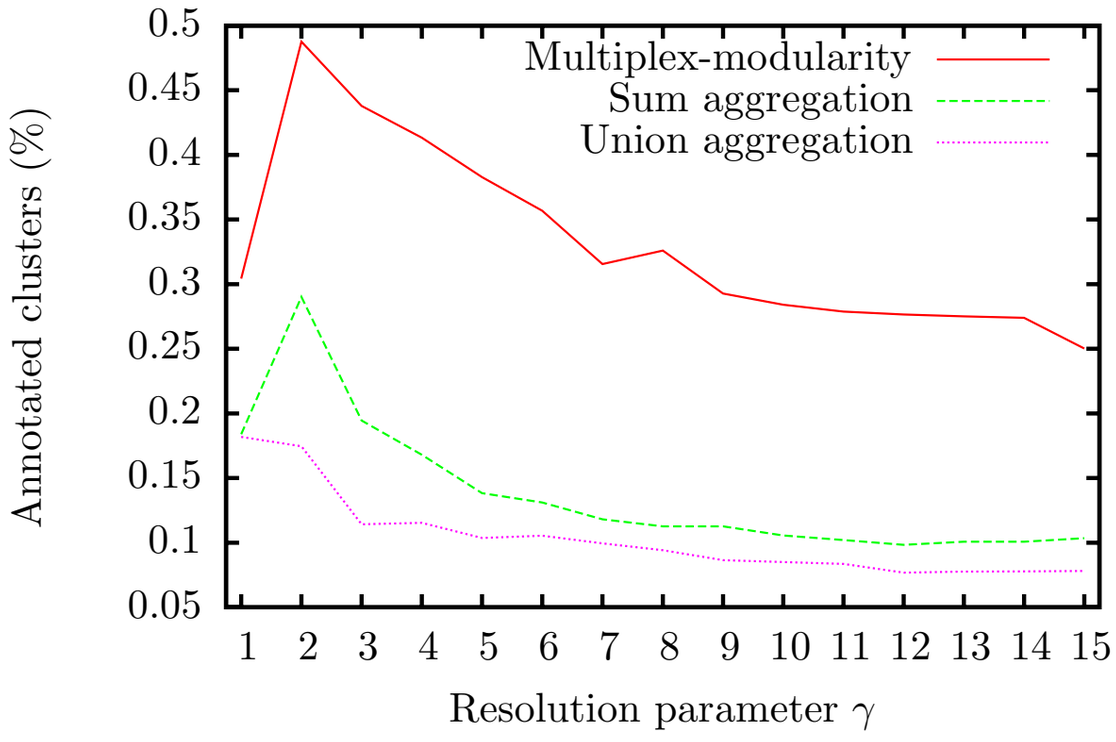




=> Considering many networks improves the detection of communities
=> Multiplex framework more efficient with mixed and incomplete networks

https://github.com/gilles-didier/**MoITi**

Real biological networks



=> More annotated

Didier et al. (2015) https://github.com/gilles-didier/**MolTi**

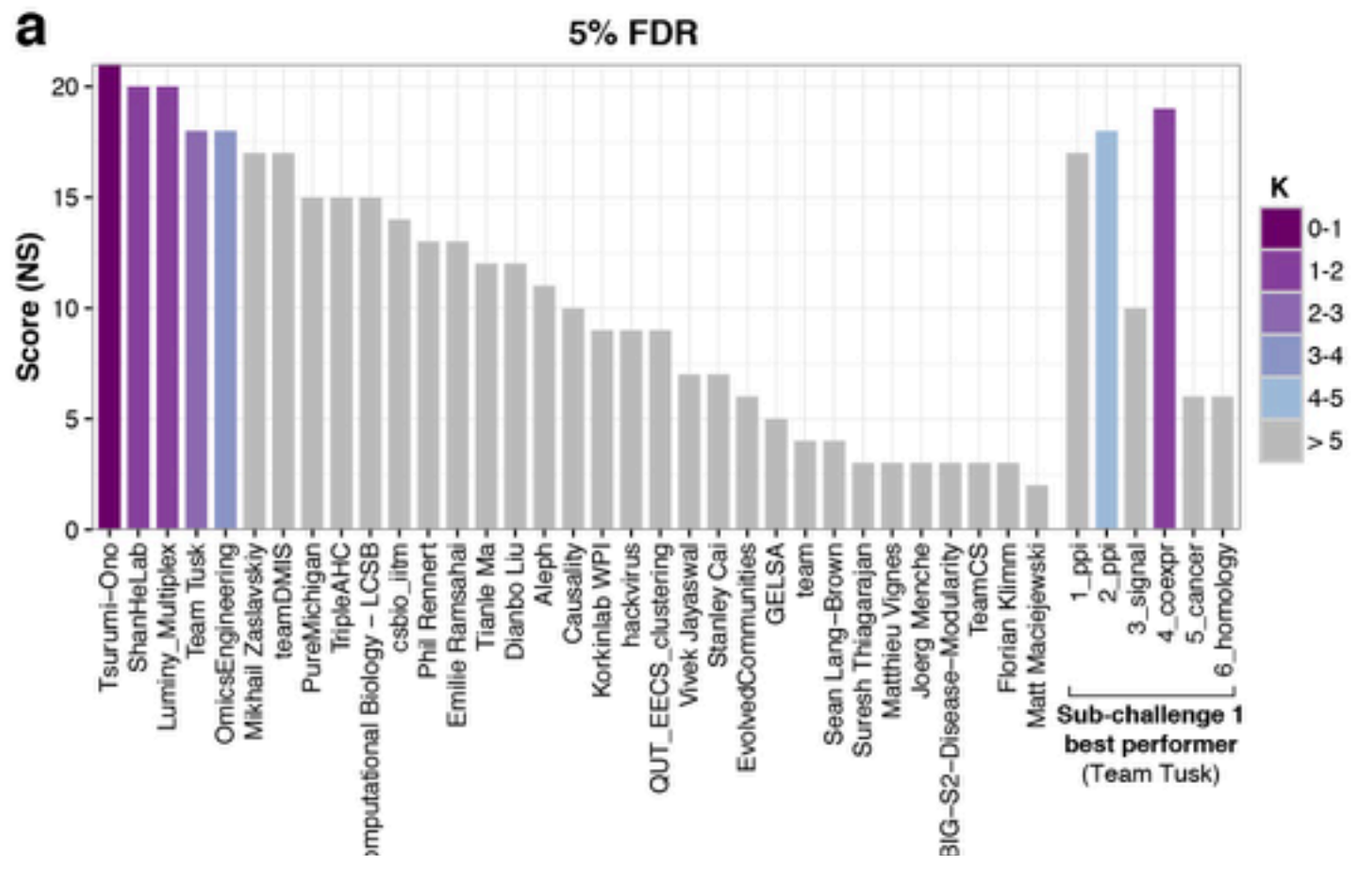
Disease Module Identification DREAM Challenge

Discover disease pathways in genomic networks

DREAM CHALLENGES powered by Sage Bionetworks

UNIL | Université de Lausanne · SIB Swiss Institute of Bioinformatics · Joint Research Center for Computational Biomedicine · RWTH AACHEN UNIVERSITY · IBM · Sage BIONETWORKS · Cell

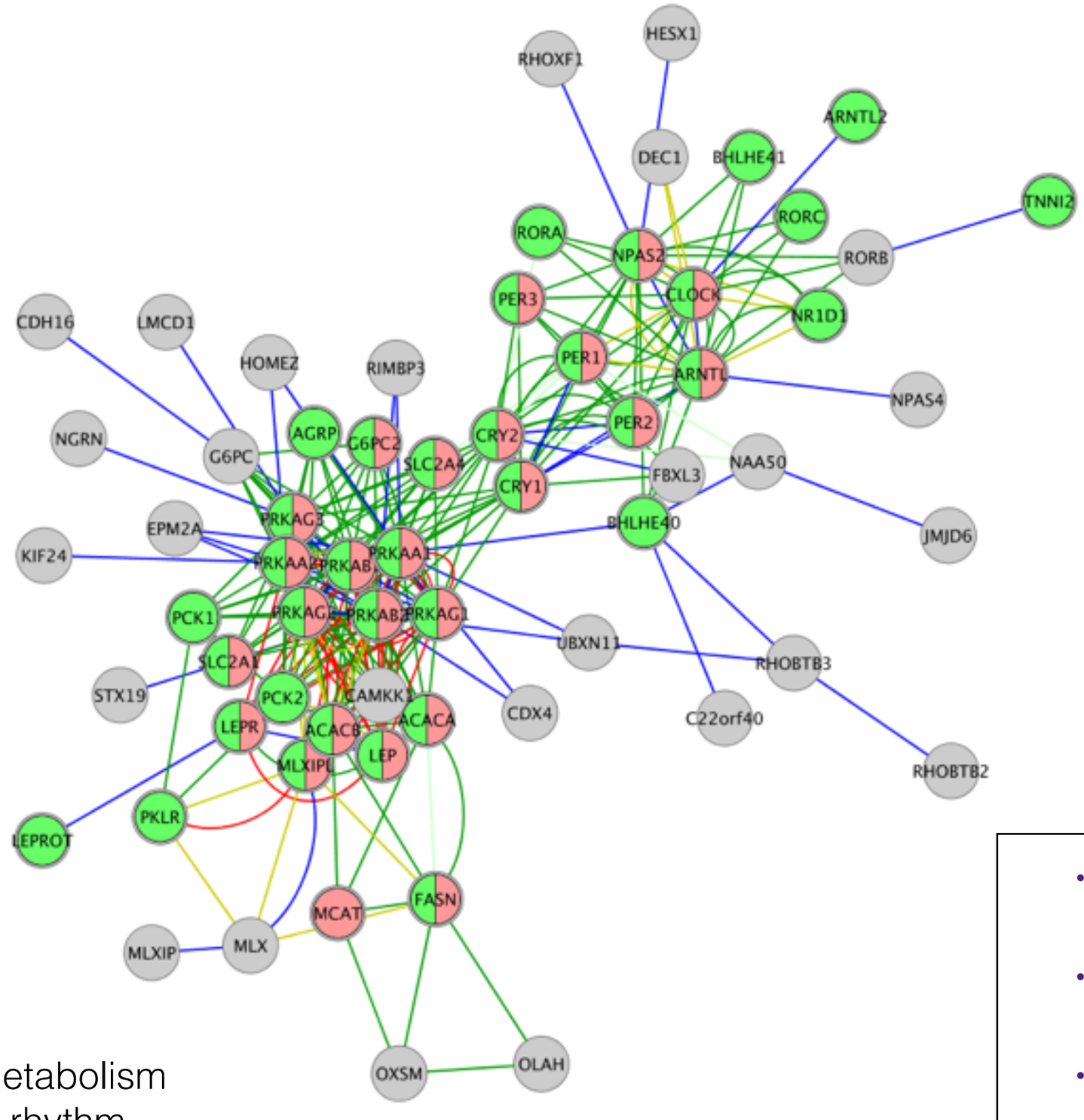
- **6 anonymous networks**
- **2 sub-challenges:**
  monoplex and multiplex
- **Modules enriched in**
  GWAS-significant genes



Molti 2.0 => randomization, recursion

Consortium paper in prep.

Disease-Community data integration



Prostate Cancer

Diabetes Mellitus

Fatty acid metabolism

Circadian rhythm

- **Processes deregulated / impacted in diseases**
- **Predict new disease-associated genes**
- **Study disease-disease relationships**

Mendelian

Polygenic

Common

Monogenic

Complex

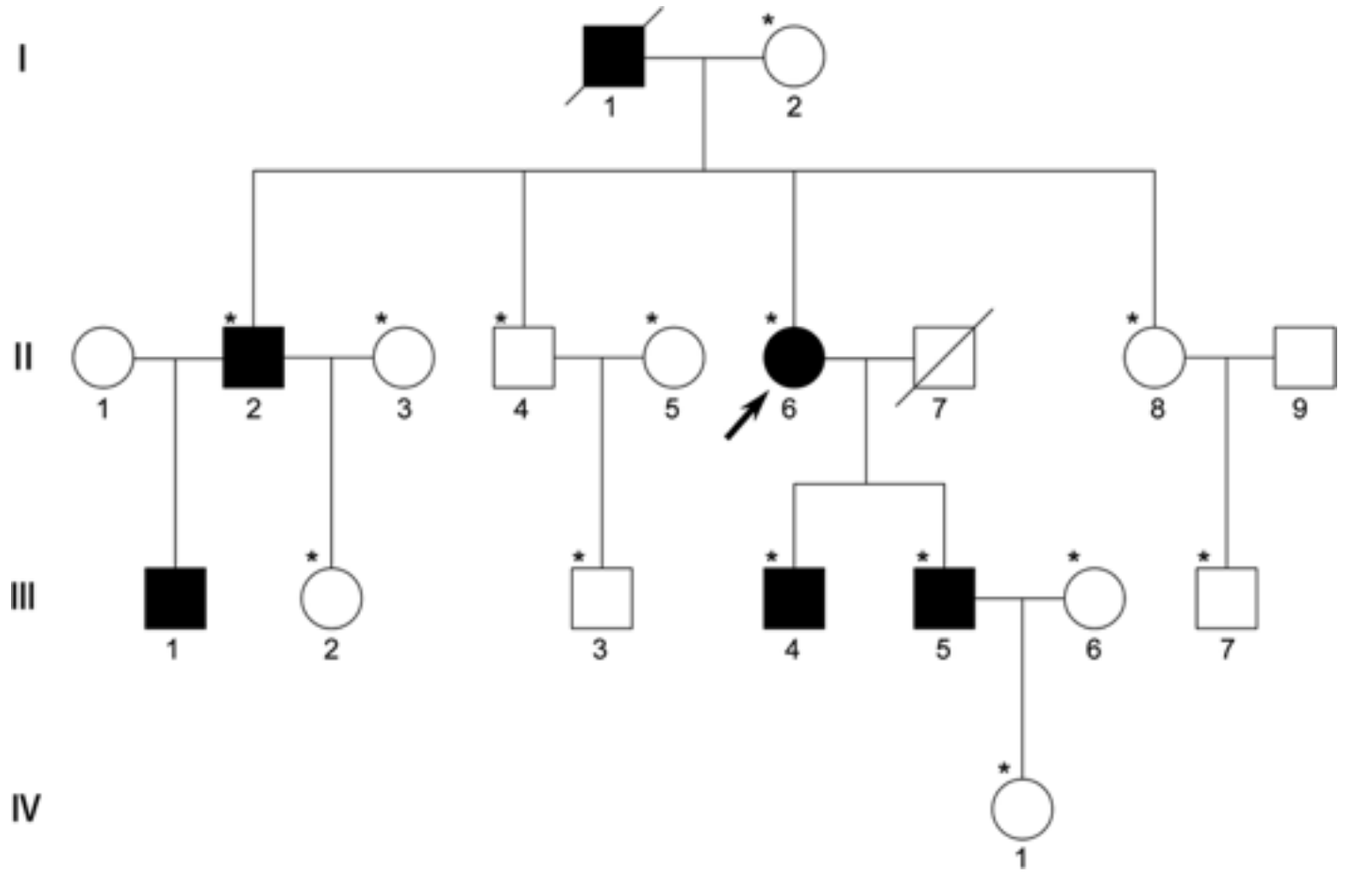
Orphan

Rare

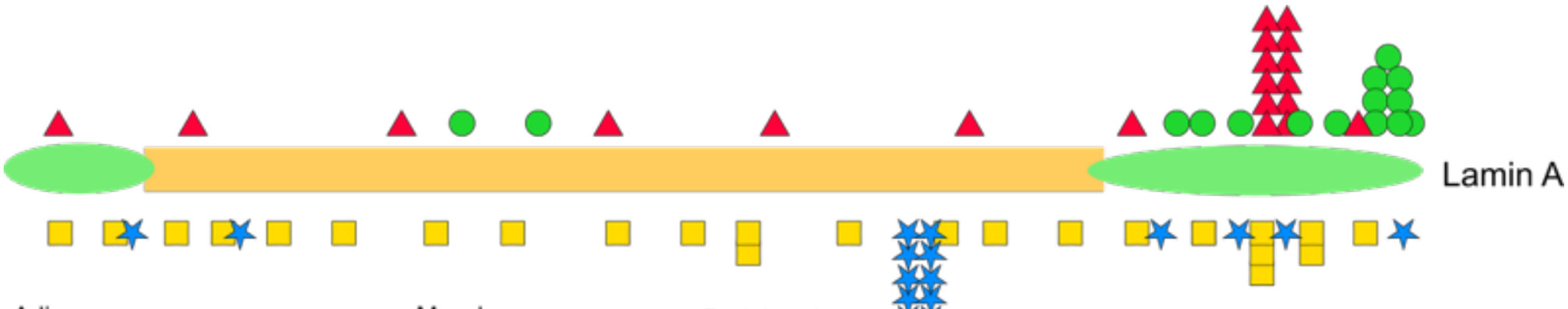
Multigenic

Rare genetic diseases

- Mono/Oligo-genics
- Mendelian inheritance / de novo
- Also very complex …



Clinical heterogeneity
Ex: Laminopathies



Adipose

▲ Familial partial Lipodystrophie, Dunnigan type

Muscle

■ Emery-Dreifuss muscular dystrophy, type 2

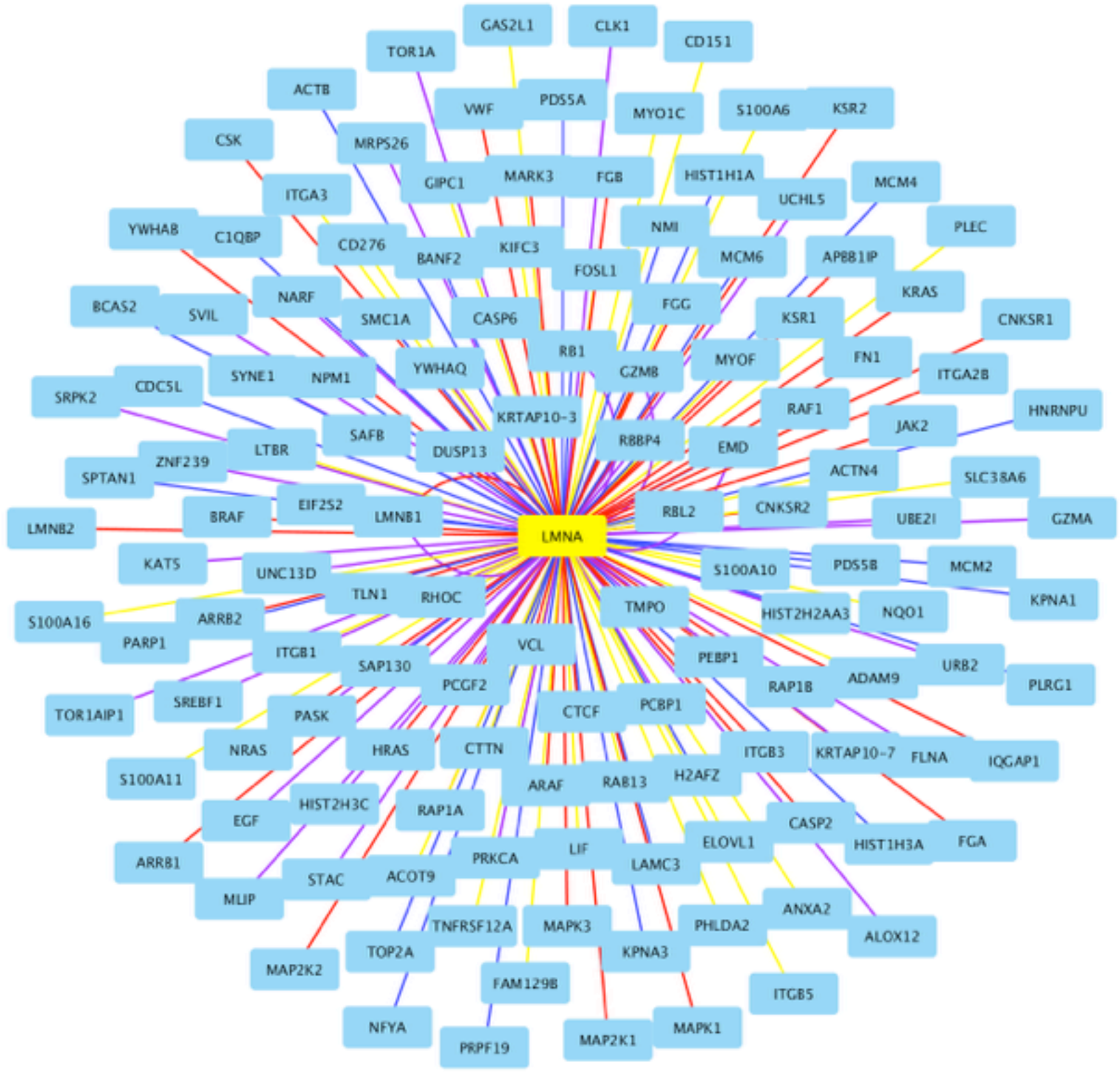
Peripheral nerve

★Charcot-Marie-Tooth, CMT2B1

Multiple

● Hutchinson-Gilford Progeria

LMNA interactome



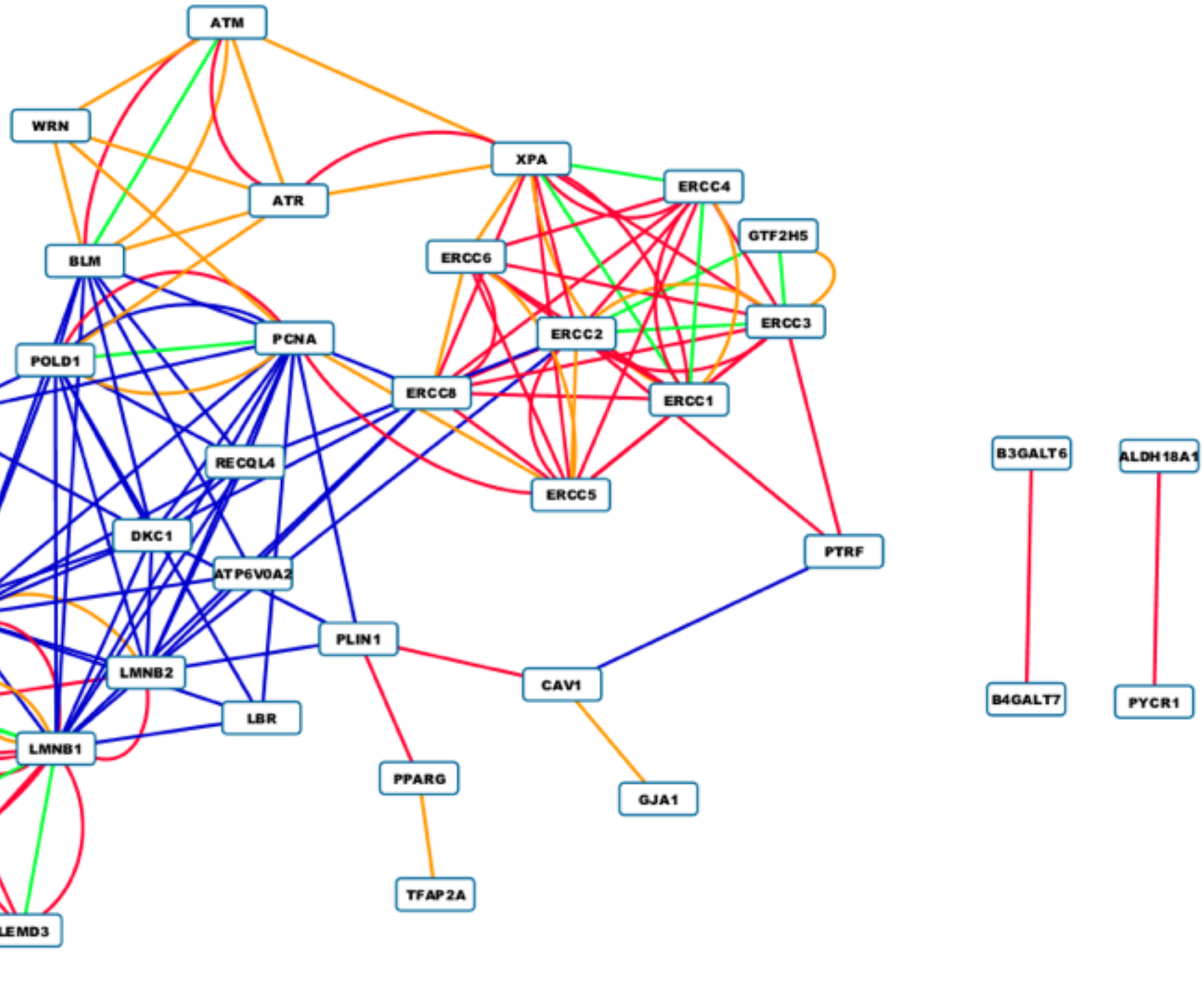
Locus heterogeneity
Ex: Premature Aging Syndromes

liste_genes_vieillissement prematuré.txt

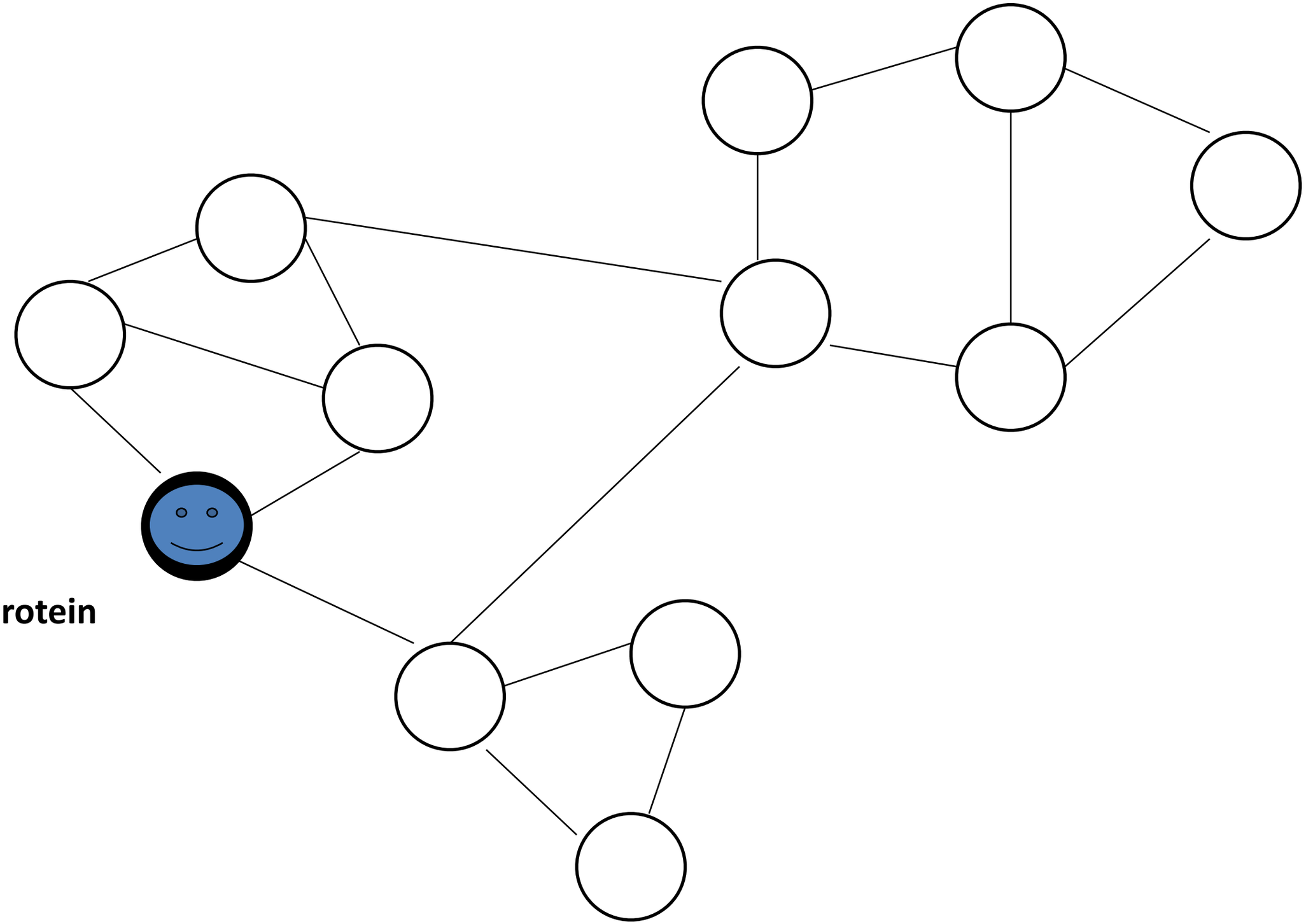
BANF1	NM_003860
BLM	NM_000057
ERCC1	NM_001983
ERCC2	NM_000400
ERCC3	NM_000122
ERCC6	NM_000124
ERCC8	NM_000082
GTF2H5	NM_207118
LMNA	NM_170707
RECQL4	NM_004260
MPLKIP	NM_138701
WRN	NM_000553
XPA	NM_000380
ERCC5	NM_000123
ZMPSTE24	NM_005857
PIK3R1	NM_181523
B4GALT7	NM_007255
B3GALT6	NM_080605
ALDH18A1	NM_002860
ATP6V0A2	NM_012463
ATM	NM_000051
ATR	NM_001184
GJA1	NM_000165
PTDSS1	NM_014754
ERCC4	NM_005236
POLD1	NM_002691
FBN1	NM_000138
TFAP2A	NM_003220
PCNA	NM_002592
TERC	NR_001566
DKC1	NM_001363
PYCR1	NM_006907

Assistance Publique
Hôpitaux de Marseille

Premature aging interactome



Random Walk with Restart

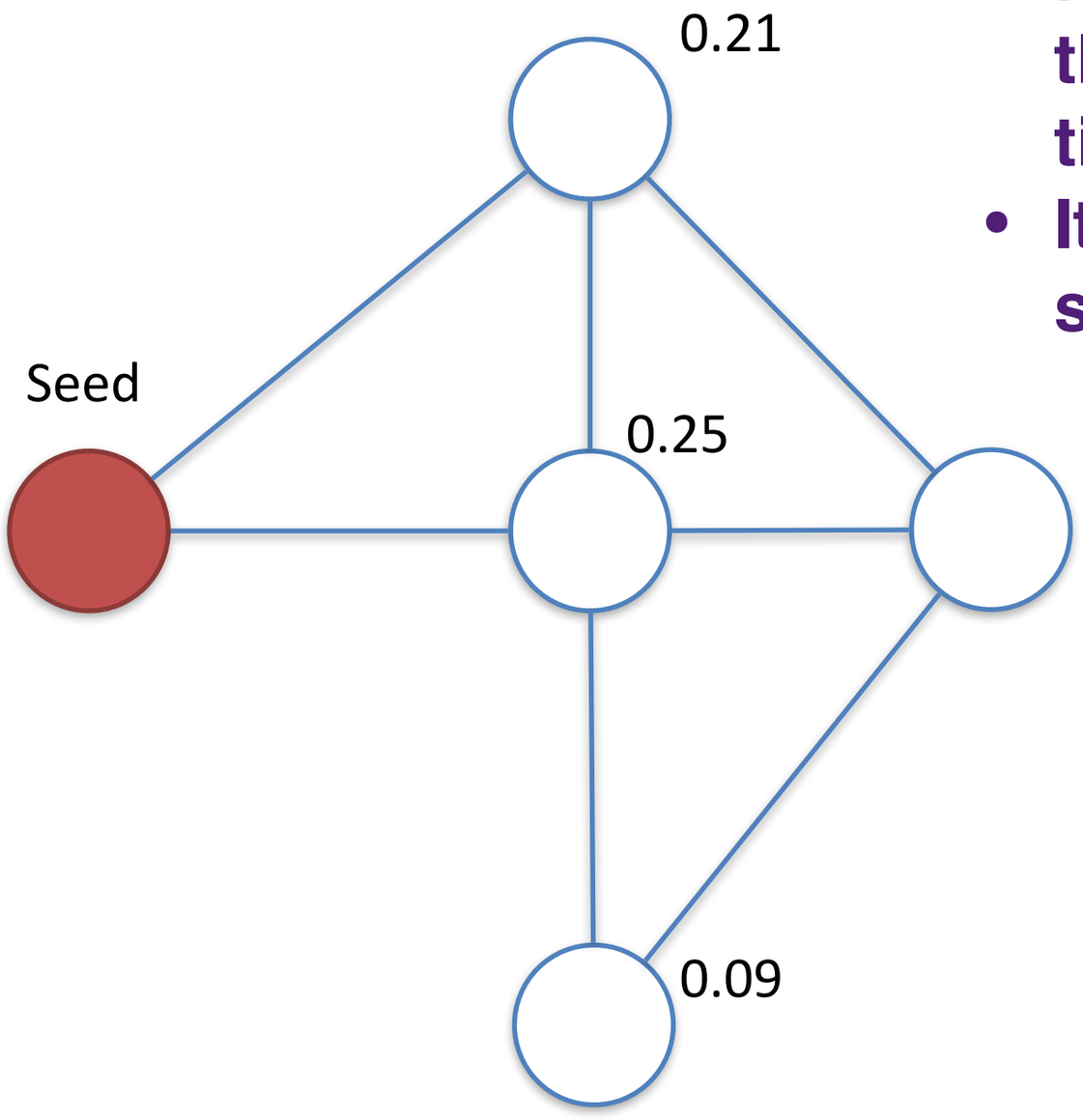


Seed Protein

Adapted from http://slideplayer.com/slide/5223771/

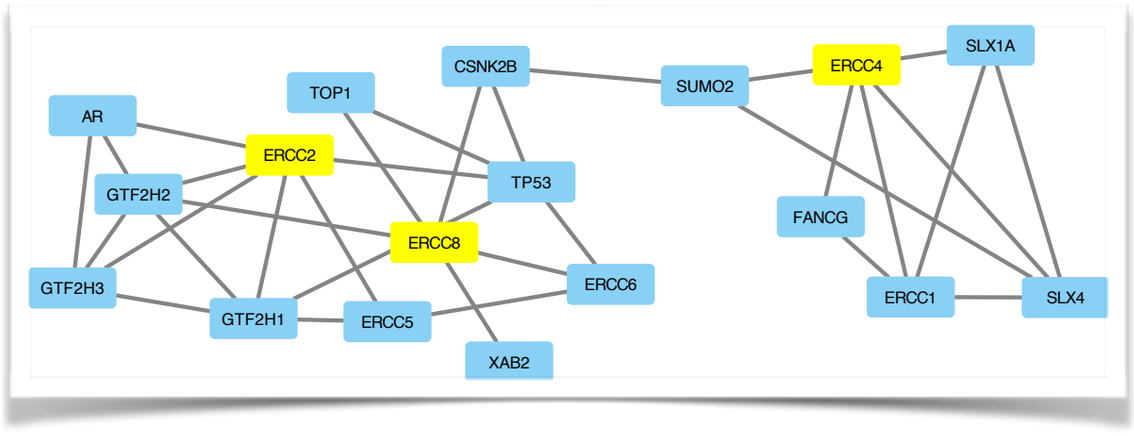
Alberto Valdeolivas

Random Walk with Restart

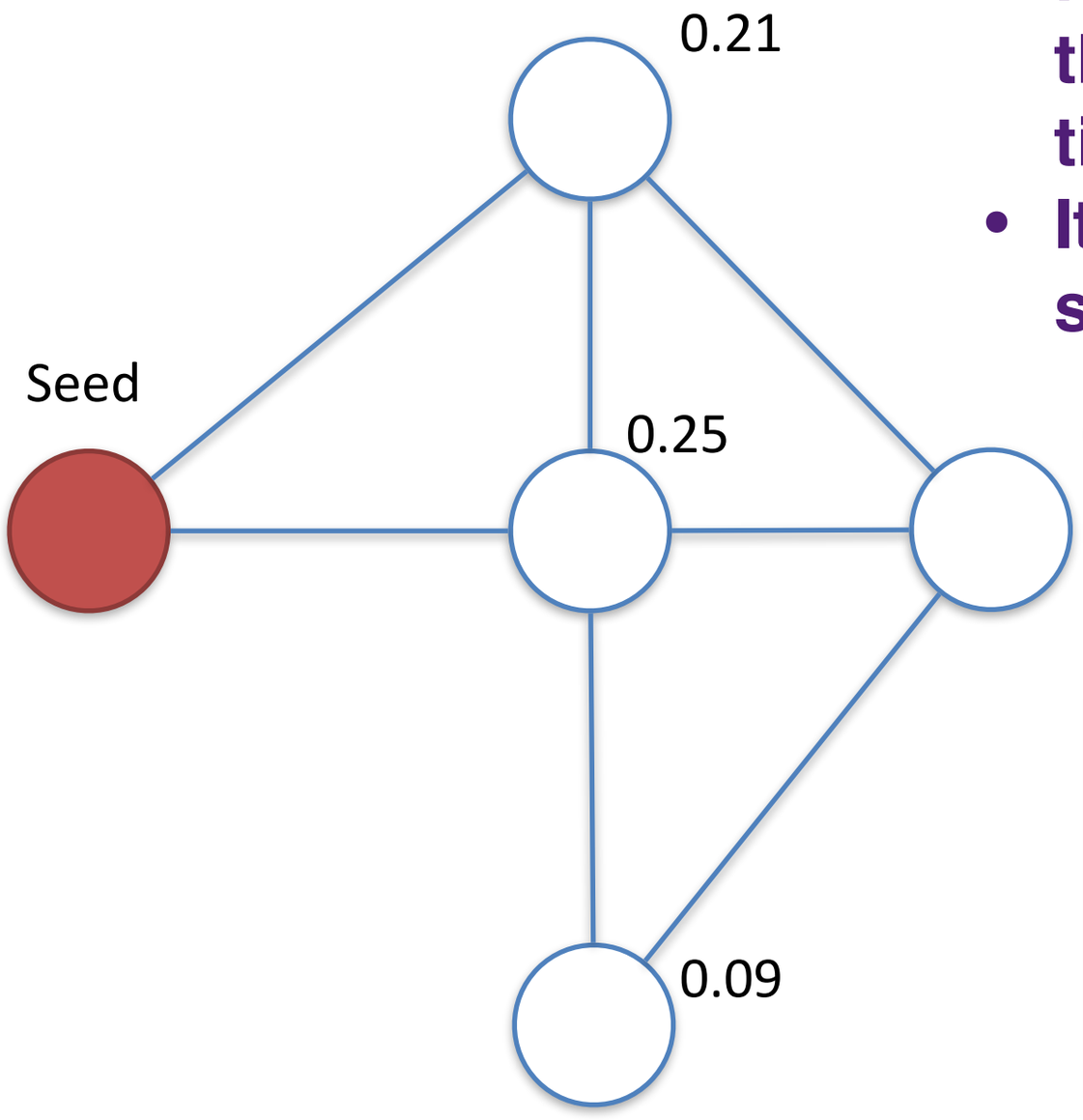


- **RWR estimates the probability that the particle visits each node when time tends to infinity**
- **It is a measure of proximity to the seeds**

$$\mathbf{p}_{t+1} = (1-r)M\mathbf{p}_t + r\mathbf{p}_0$$

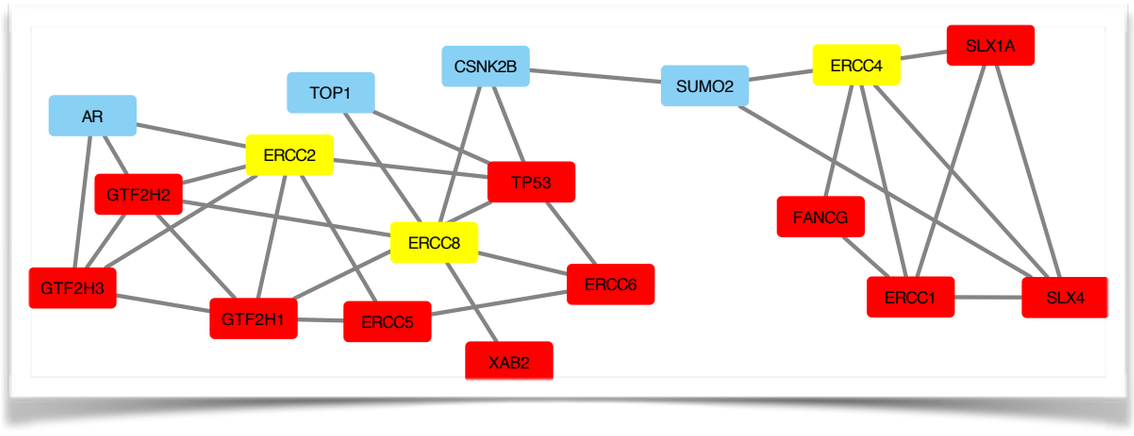


Random Walk with Restart

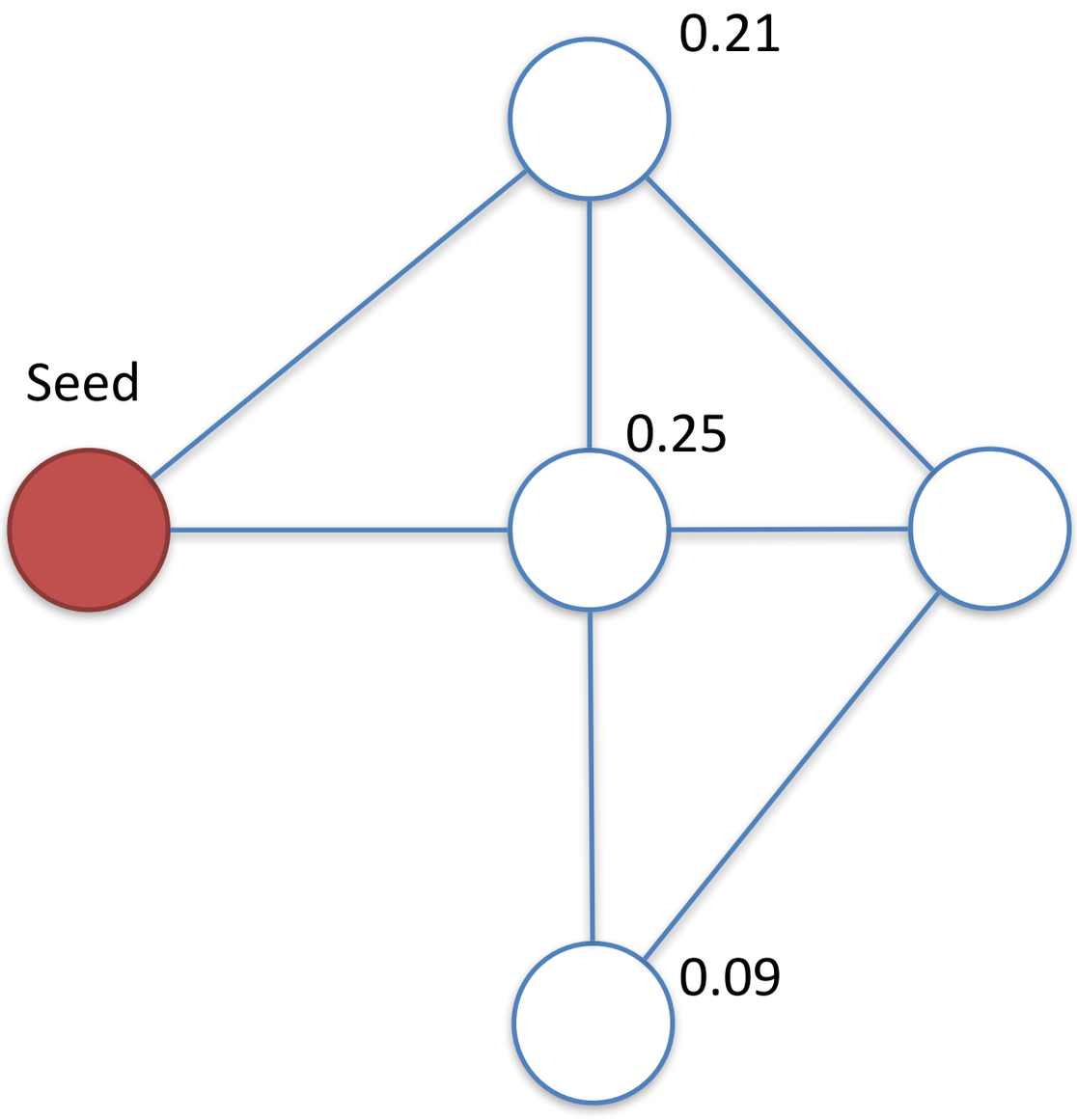


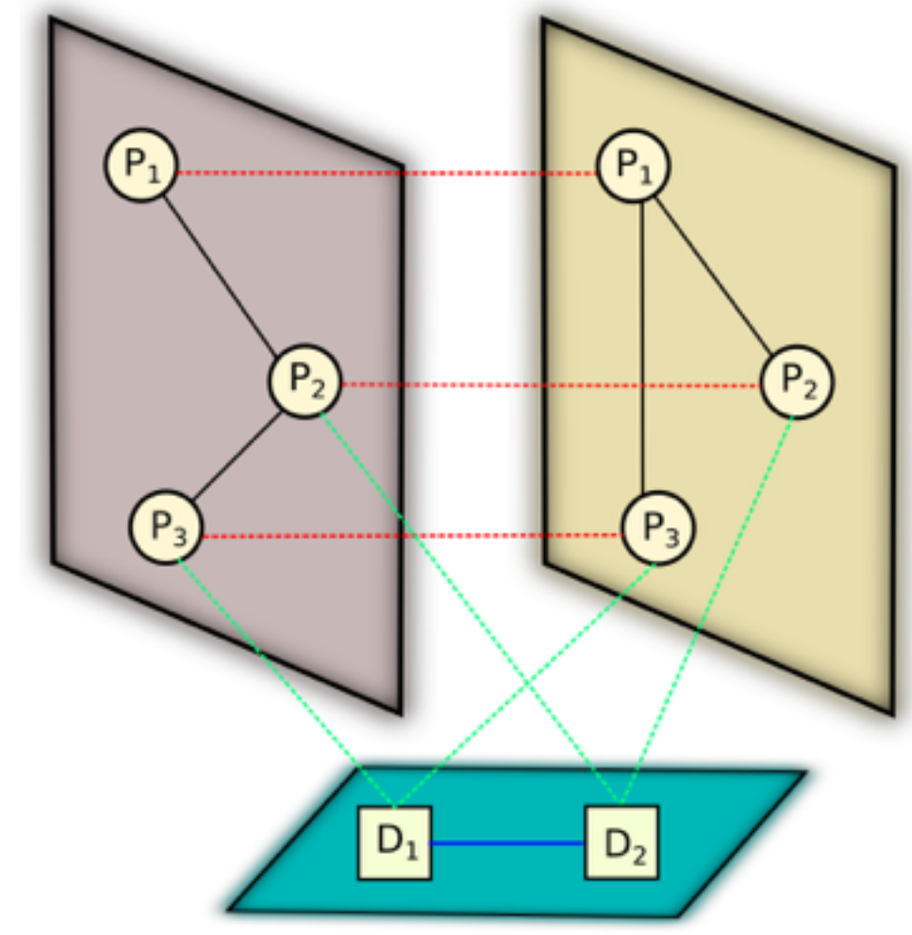
- **RWR estimates the probability that the particle visits each node when time tends to infinity**
- **It is a measure of proximity to the seeds**

$$\mathbf{p}_{t+1} = (1-r)M\mathbf{p}_t + r\mathbf{p}_0$$



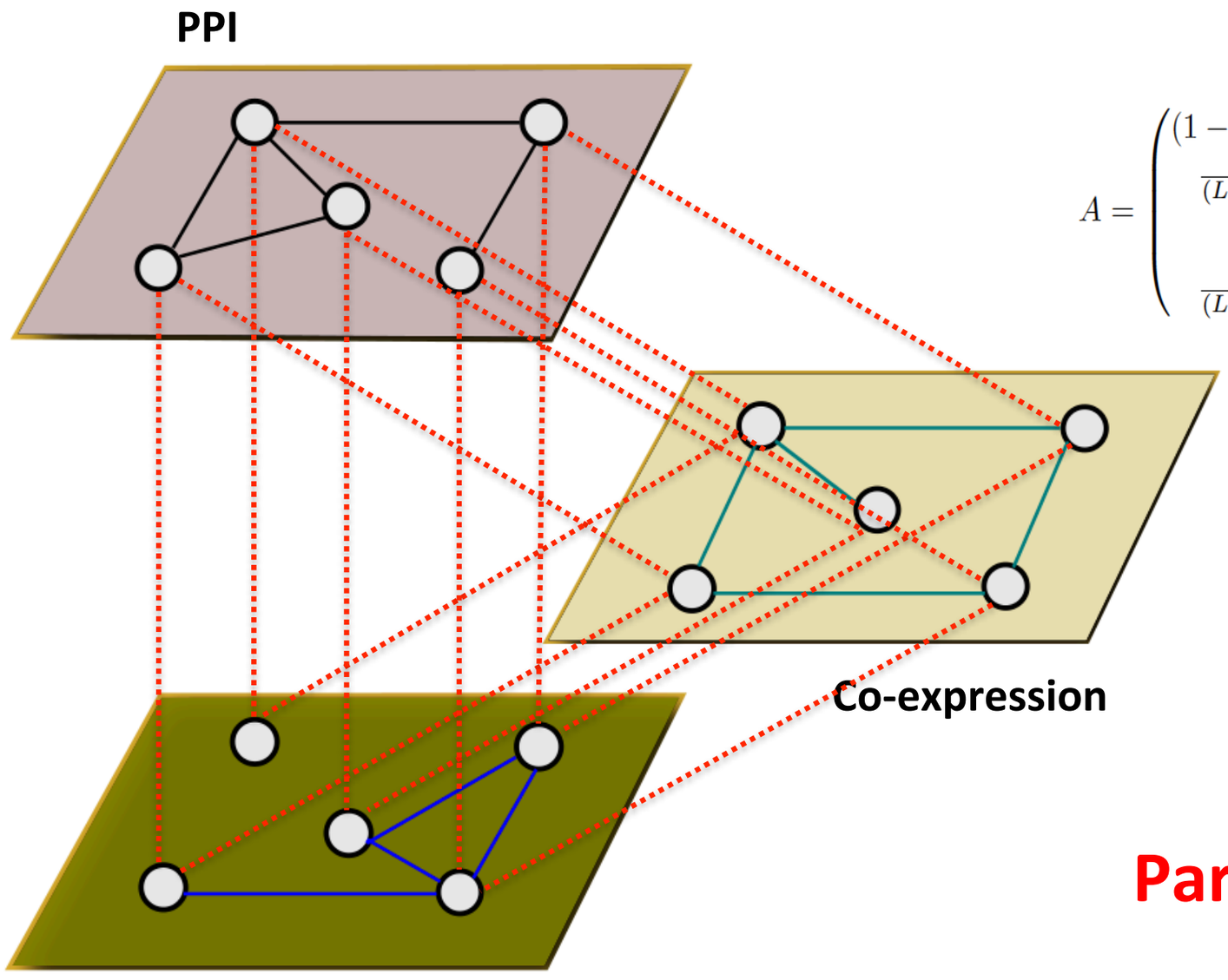
Random Walk with Restart in Multiplex and Heterogeneous Networks





Alberto Valdeolivas

Extension of RWR to Multiplex Networks (RWR-M)

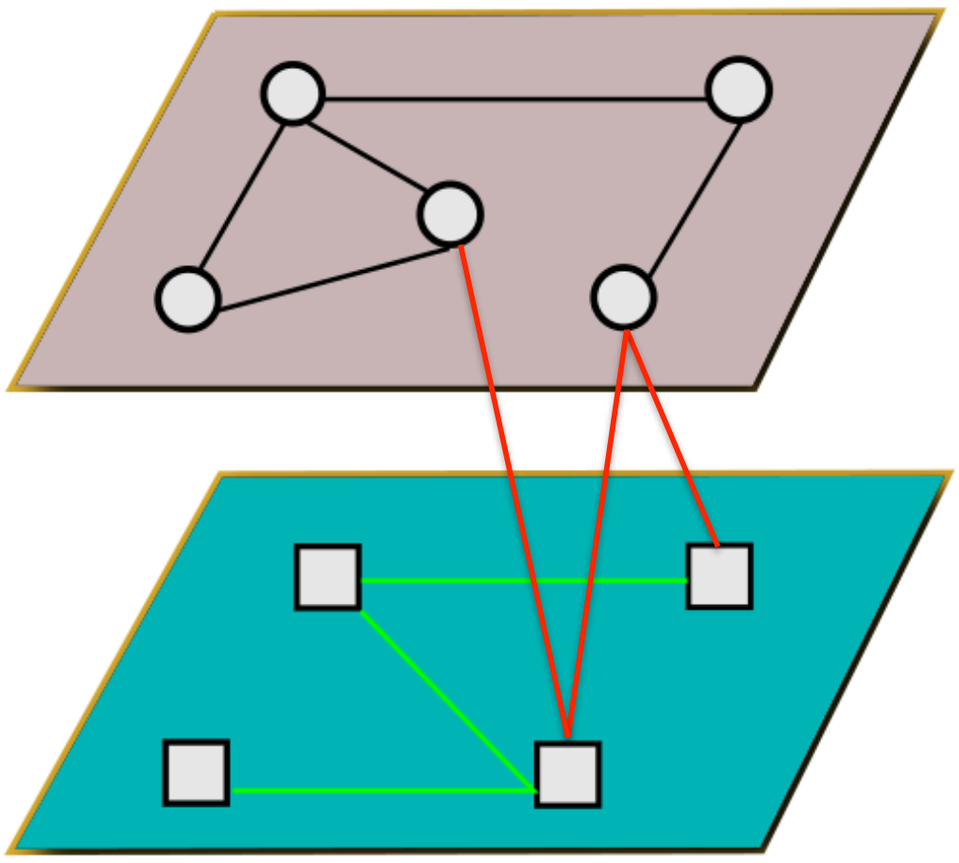


$$A = \begin{pmatrix} (1-\delta)A^{[1]} & \frac{\delta}{(L-1)}\mathbf{I} & \cdots & \frac{\delta}{(L-1)}\mathbf{I} \\ \frac{\delta}{(L-1)}\mathbf{I} & (1-\delta)A^{[2]} & \cdots & \frac{\delta}{(L-1)}\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta}{(L-1)}\mathbf{I} & \frac{\delta}{(L-1)}\mathbf{I} & \cdots & (1-\delta)A^{[L]} \end{pmatrix}$$

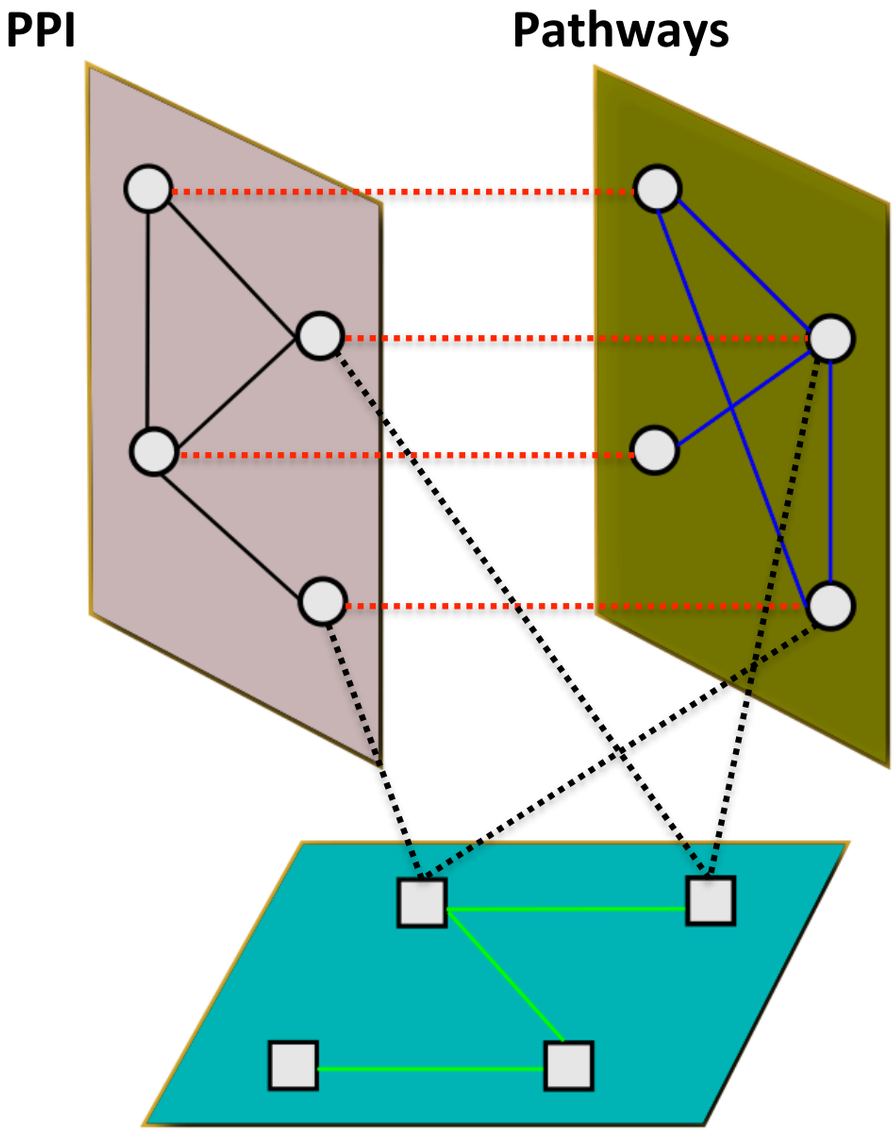
$$\overline{\mathbf{p}}_{t+1} = (1-r)M\overline{\mathbf{p}}_t + r\overline{\mathbf{p}}_{RS}$$

Parameters to tune:
δ and τ

Heterogeneous Networks



Extension of RWR to Multiplex and Heterogeneous Networks (RWR-MH)



$$\widetilde{\mathbf{p}}_{t+1} = (1-r)H\widetilde{\mathbf{p}}_t + r\widetilde{\mathbf{p}}_{RS}$$

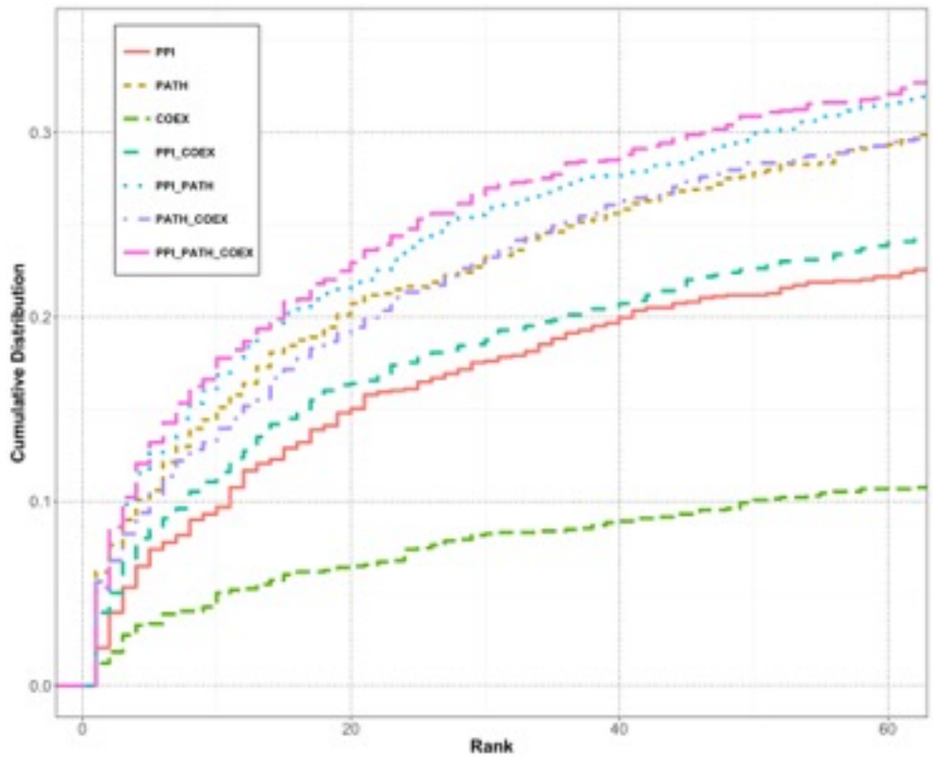
$$H = \begin{bmatrix} H_{PP} & H_{PD} \\ H_{DP} & H_{DD} \end{bmatrix}$$

$$H_{DP}(a,j) = \begin{cases} \lambda B(j,a)/\sum_{k=1}^{n} B(k,a), & \text{if } \sum_{k=1}^{n} B(k,a) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

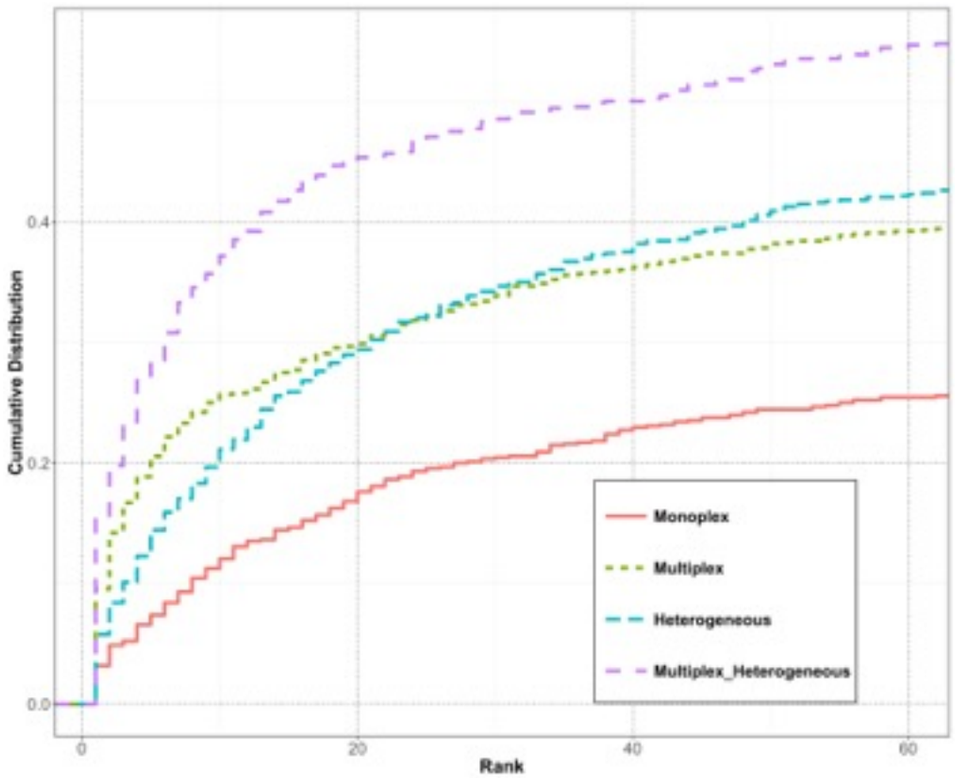
$$H_{DD}(a,b) = \begin{cases} A_D(a,b)/\sum_{k=1}^{m} A_D(a,k), & \text{if } \sum_{k=1}^{n} B(k,b) = 0 \\ (1-\lambda)A_D(a,b)/\sum_{k=1}^{m} A_D(a,k), & \text{otherwise} \end{cases}$$

Parameters to tune:
λ and η

Leave-One-Out Cross Validation



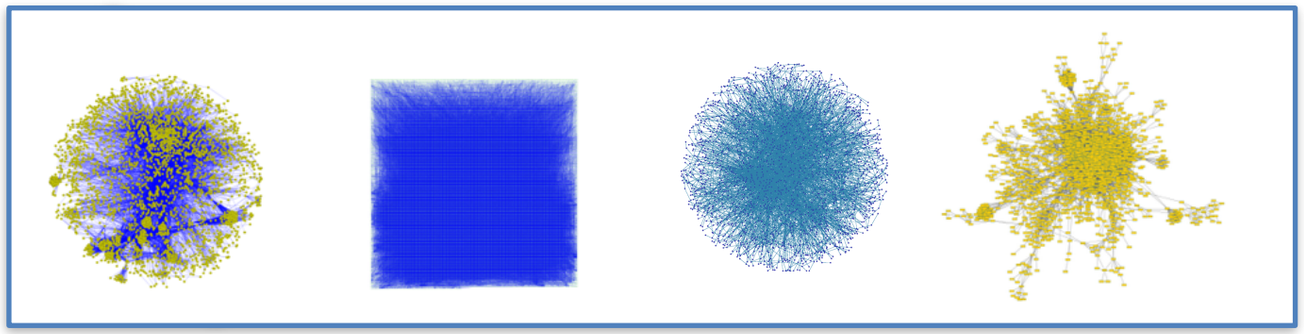
Multiplex versus monoplex



Multiplex-heterogeneous

Valdeolivas et al. bioRxiv 2017.

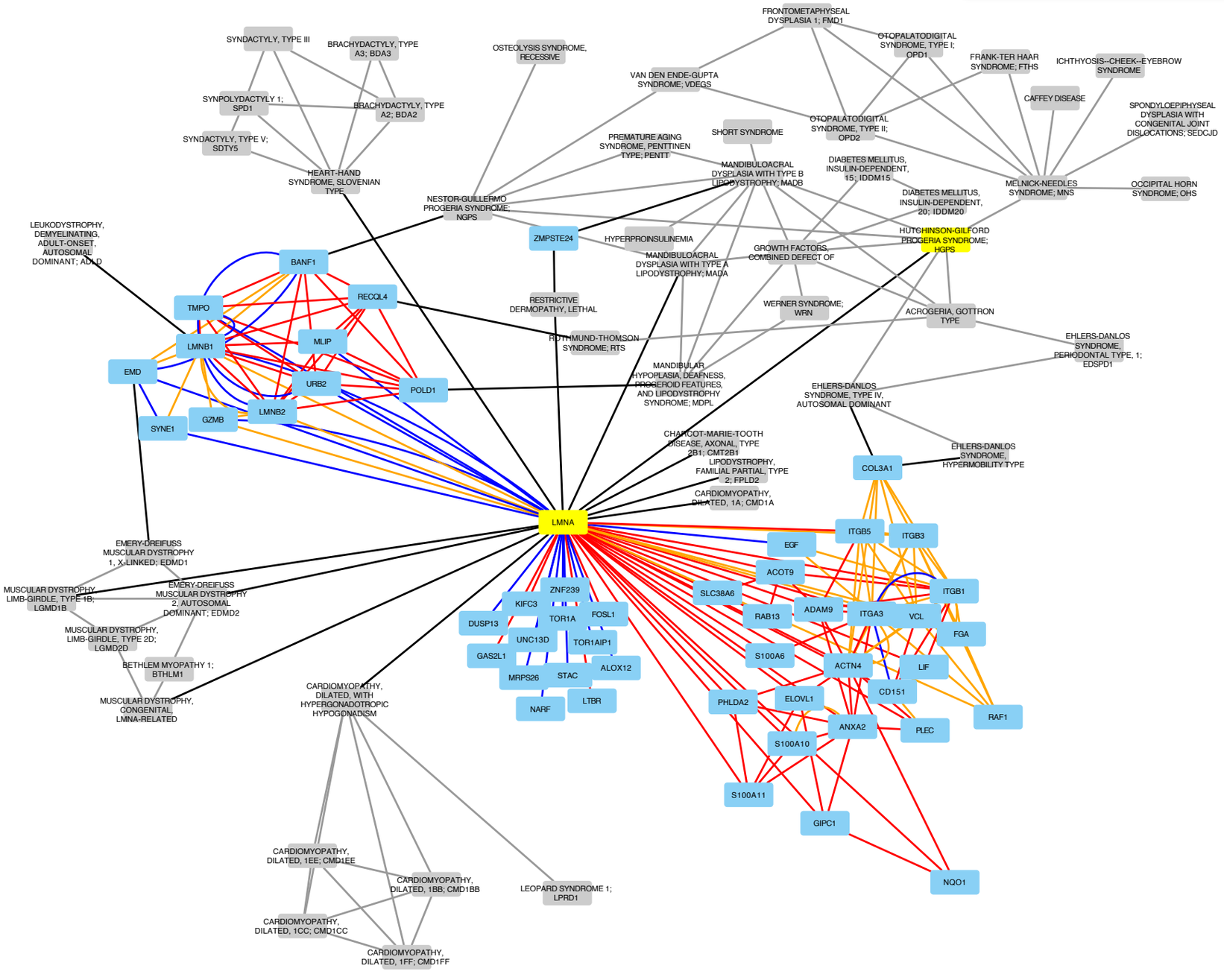
Application: Explore Network vicinity of disease and disease gene

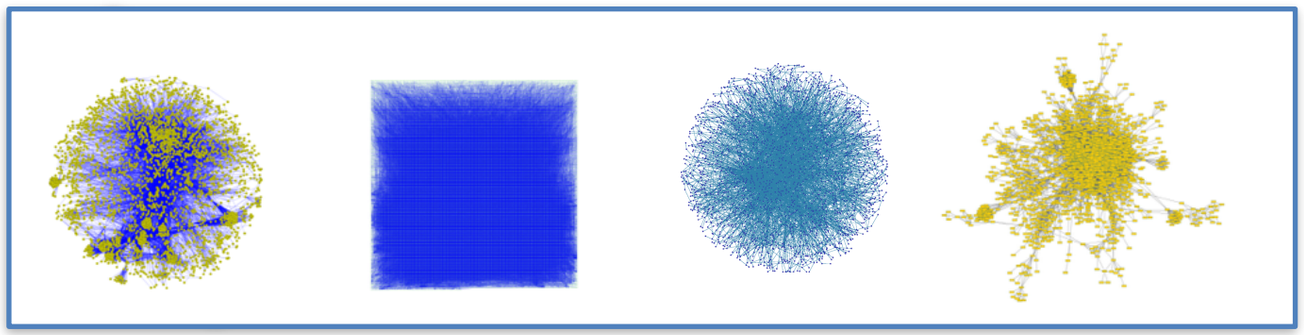


Disease-disease network

Random Walk with Restart

Top 50

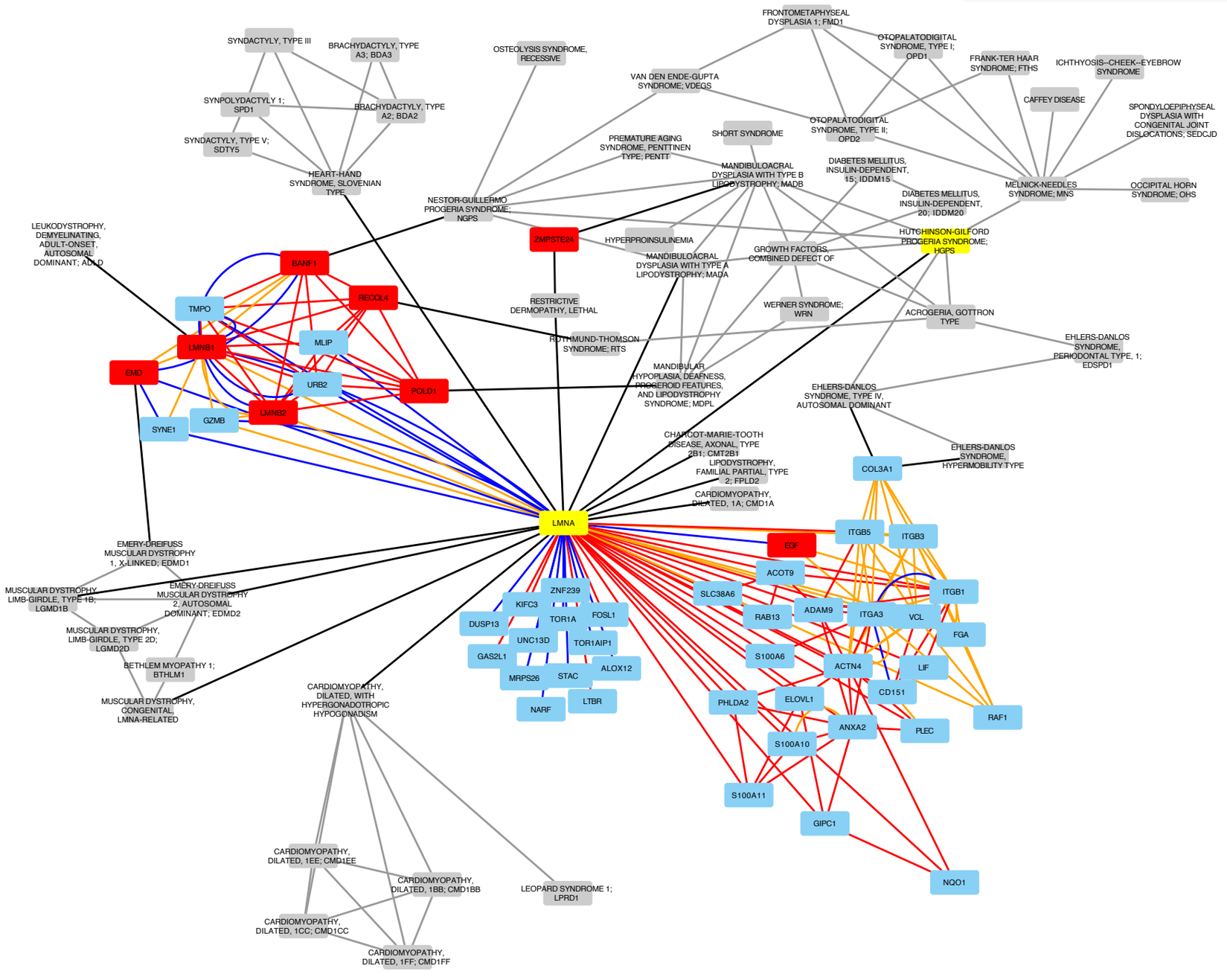




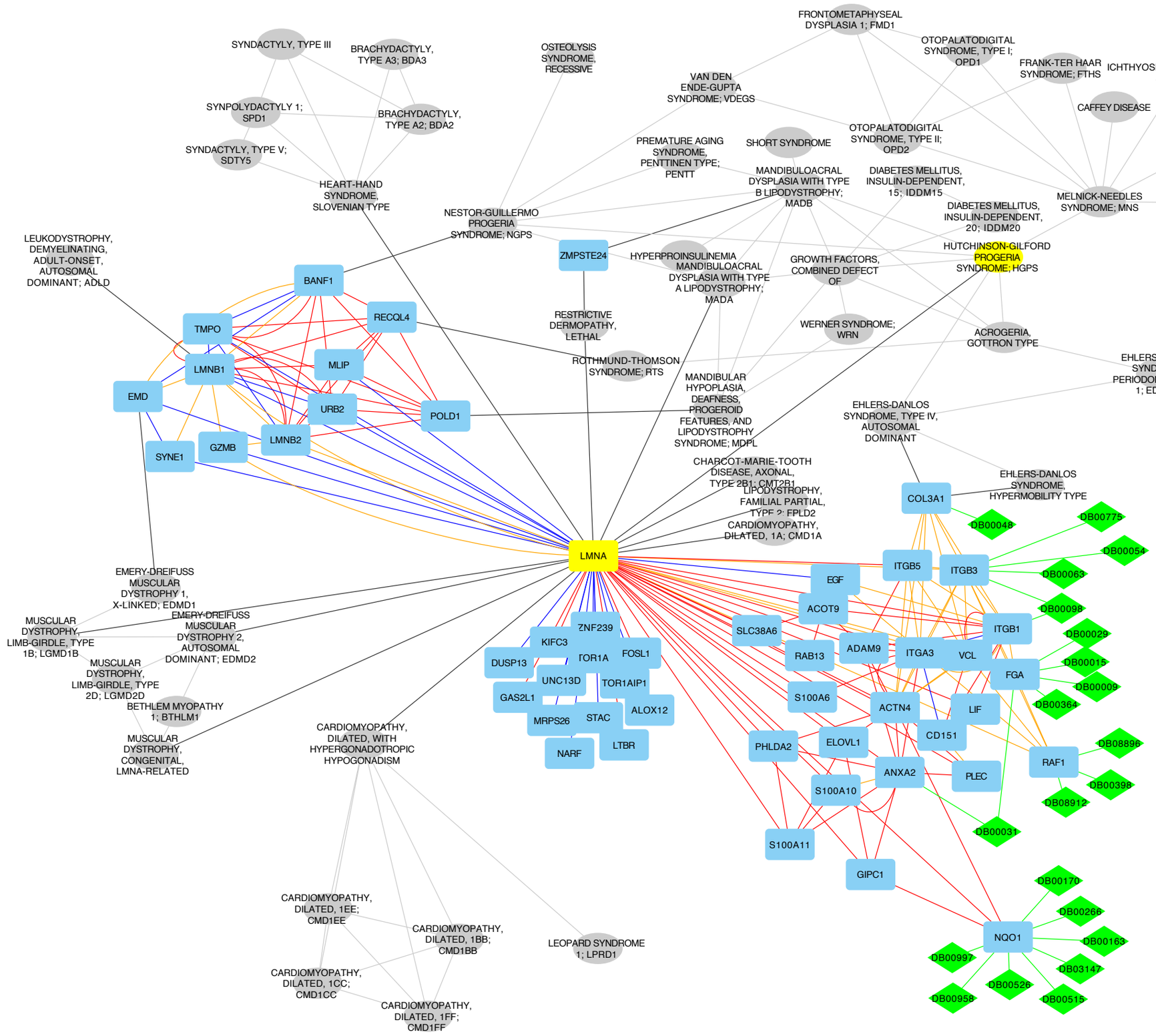
Application: Explore Network vicinity of disease and disease gene

Disease-disease network

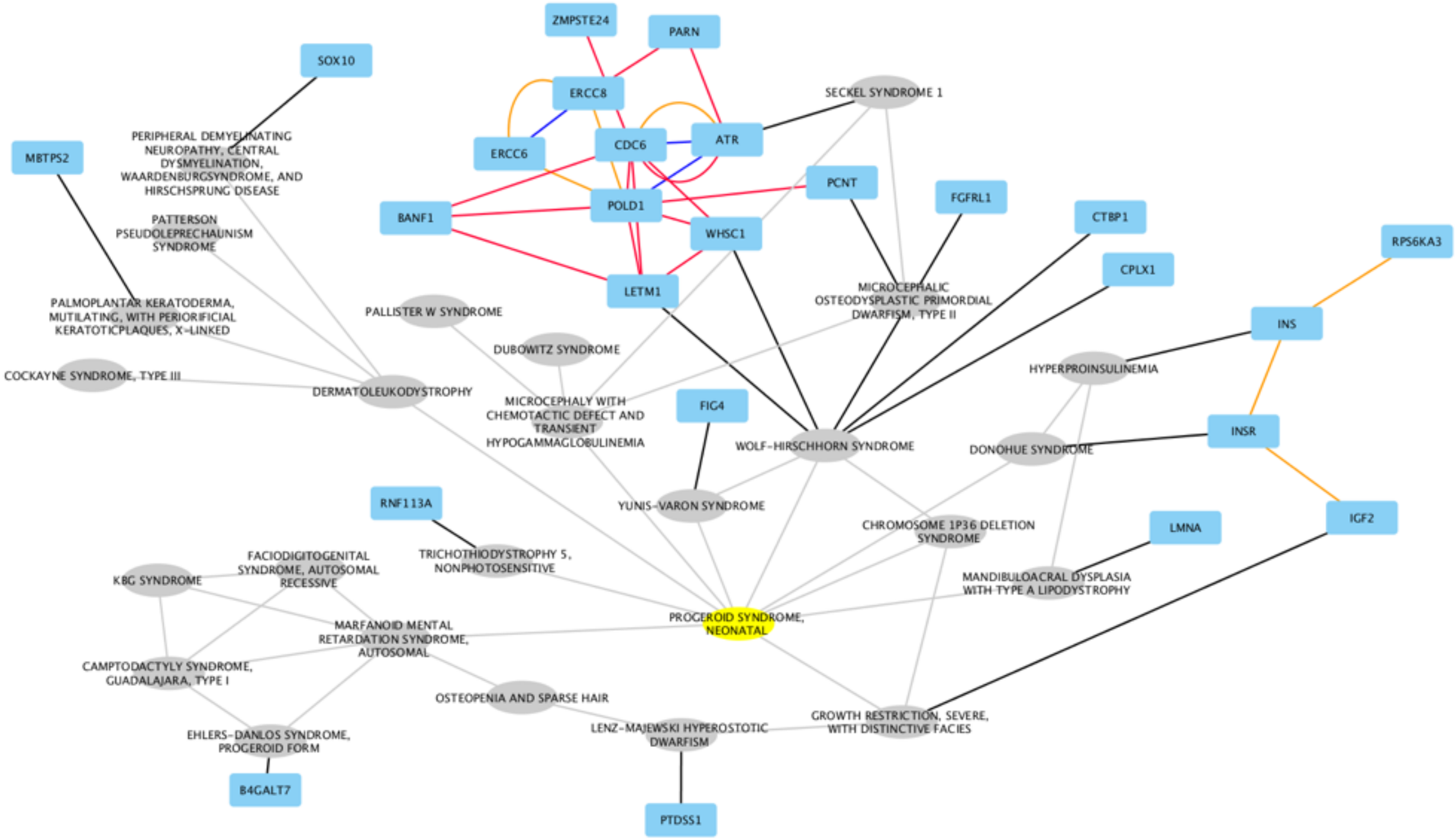
Physiological aging



Application: drug repurposing

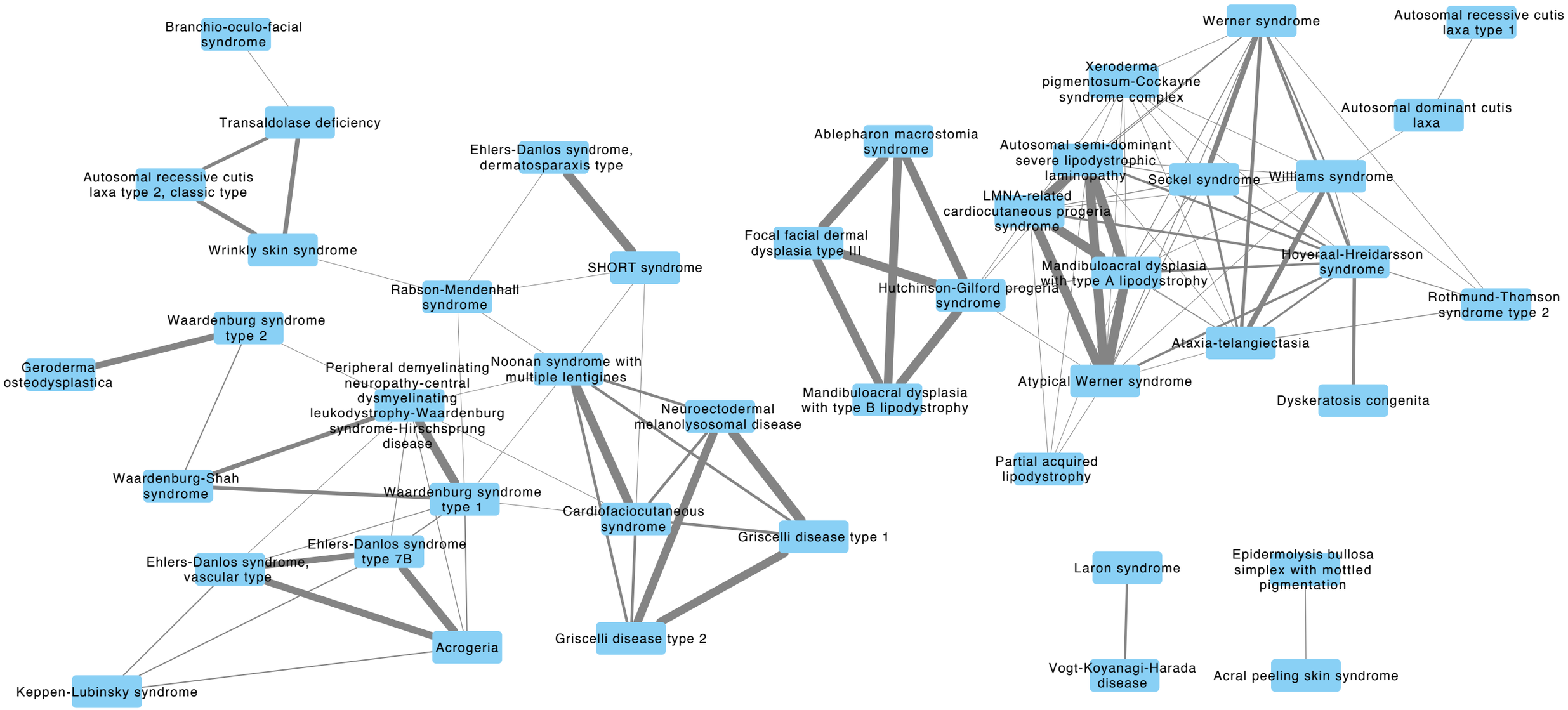


Application: diagnosis NGS



Application: disease classification and comorbidities

- All ~50 diseases with “premature aging” phenotype in HPO => associated genes
- For each disease => Top 25 genes - subnetworks with RWR-M



Plan cancer

PEPS Bio-Maths-Info

PROGELIFE
Innovative research solutions for healthy aging

A*Midex
Initiative d'excellence Aix-Marseille

Alberto Valdeolivas
Elva Novoa

Patrice Roll
Annachiara De Sandre

Nicolas Lévy

+ PostDocs !!!

Pierre Cau
Claire Navarro
Sophie Perrin

INSTITUT DE MATHÉMATIQUES de MARSEILLE

Elisabeth Remy
Laurent Tichit
Alain Guénoche
Brigitte Mossé
Gilles Didier

PROGELIFE
Innovative research solutions for healthy aging

JOBIM 2018

03>06 JUIL

PALAIS DU PHARO | MARSEILLE

JOURNÉES OUVERTES DE BIOLOGIE INFORMATIQUE & MATHÉMATIQUES

Marseille
Medical
Genetics