

Биолошке мреже са фокусом на интерактомику

Лазар Васовић

Математички факултет, Универзитет у Београду
<https://github.com/matfija/Neuredjenost-u-interaktomu>

мај 2022



Универзитет у Београду
Математички факултет

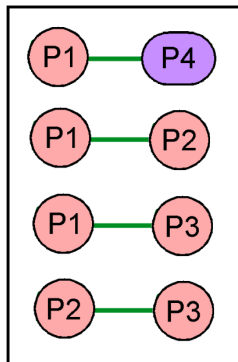
- 1 Биолошке мреже
- 2 Мреже протеинских интеракција
- 3 Неуређеност у интерактому

Биолошке мреже

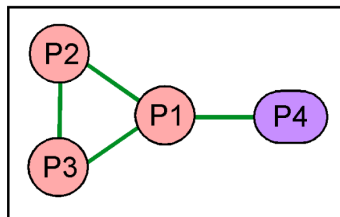
- Биолошки системи, било да је реч о маленој ћелији или огромном екосистему, састоје се од основних градивних јединица, нпр. гена, протеина, метаболита или јединки, које међудејствују.
- Ретке су јединице које делују у потпуности самостално, тако да су интеракције кључни део сваког биолошког система. Заправо се ниједан систем не може посматрати као прост збир елемената.
- За потребе биоинформатичке анализе, неопходно је одабрати одговарајући математички модел. Природни модел интерагујућих блокова, тиме и биолошких система, јесте граф (мрежа), уз напомену да он ипак не успева да представи све односе.

Моделовање биолошког система графом

ЗНАЊЕ



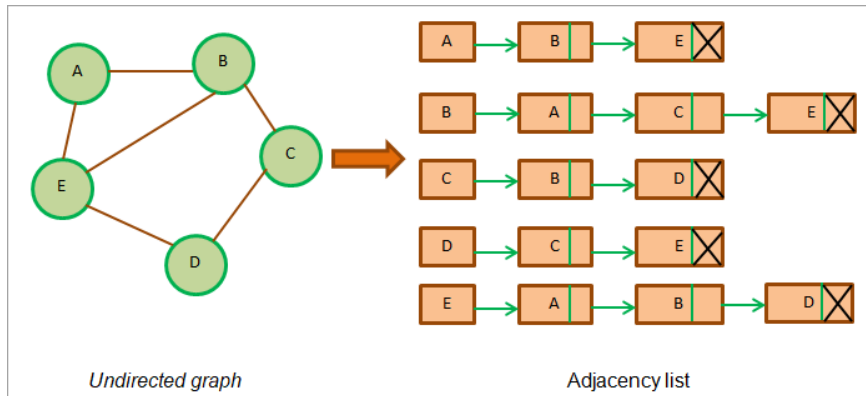
МОДЕЛ



Теорија графова

- Граф G је уређени пар коначних скупова чворова V (*vertices*) и грана (ивица, веза) E (*edges*). Математички записано, граф G је $G = (V, E)$, скуп V од n чворова $V = \{v_1, \dots, v_n\}$, а скуп E од m грана $E = \{e_1, \dots, e_m\}$. Грана $e_k = (v_i, v_j)$ повезује чвор v_i са v_j .
- Уколико није важан смер везе, гране су неуређени парови чворова, а граф је неусмерен. Пример везе: протеинска интеракција.
- Уколико јесте важан смер везе, гране су уређени парови чворова, а граф је усмерен (диграф). Пример везе: генски утицај.
- Остали важни појмови: пут, циклус, компоненте повезаности, дрво (стабло), мултиграф, графовски алгоритми (нпр. обилазак)...

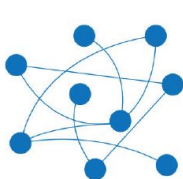
Једноставан пример графа



Особине биолошких мрежа

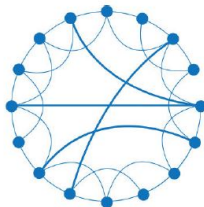
- Мере централности претпостављају значај чвора у мрежи.
 - Степен чвора (централност степена) – број суседа у мрежи.
- Тополошке карактеристике графа описују целокупну мрежу.
 - Пречник (*diameter*, D) – максимална дужина најкраћег пута.
 - Карактеристична дужина пута (*average path length*, L) – просечна дужина најкраћих путева између свака два чвора.
 - Коефицијент груписања (*clustering coefficient*, CC) – просечна склоност чворова да граде густо повезане групе (кластере).
 - Расподела степена (*degree distribution*, P_k) – вероватноћа k .
- Очекују се мање вредности D и L („мали свет”), као и велики CC (јака структура заједнице). Расподела чворова је углавном степена (*power law* или *scale-free*, $P_k \sim k^{-\gamma}$), према којој већина чворова има мали број суседа, уз мањи број високоповезаних хабова.

Сложенији примери графова



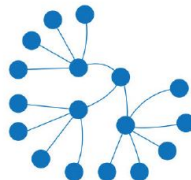
Random

Average distributions.
No structure or hierarchal patterns.



Small-World

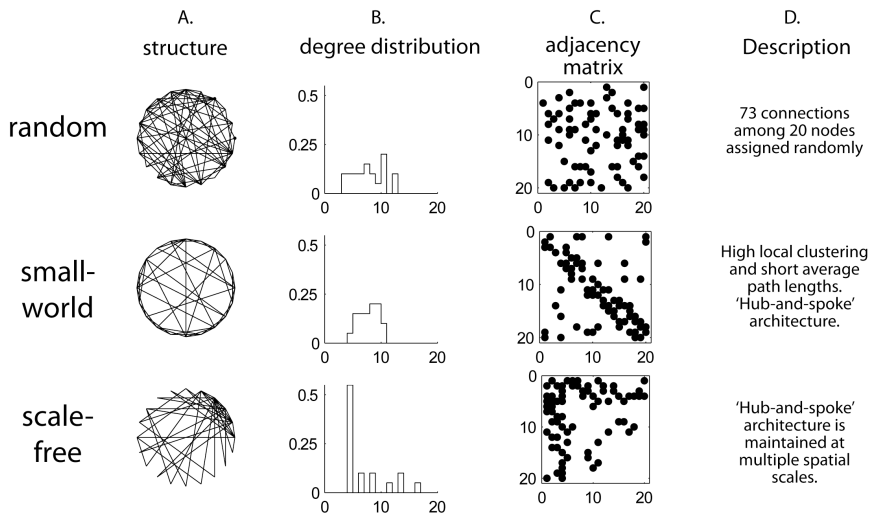
High local clustering and short
average path lengths.
Hub-and-spoke architecture.



Scale-Free

Hub-and-spoke architecture preserved at
multiple scales.
High power law distribution.

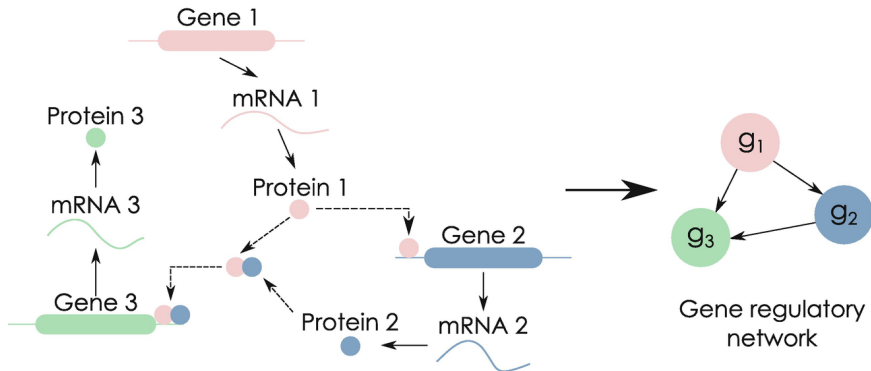
Сложеније поређење графова



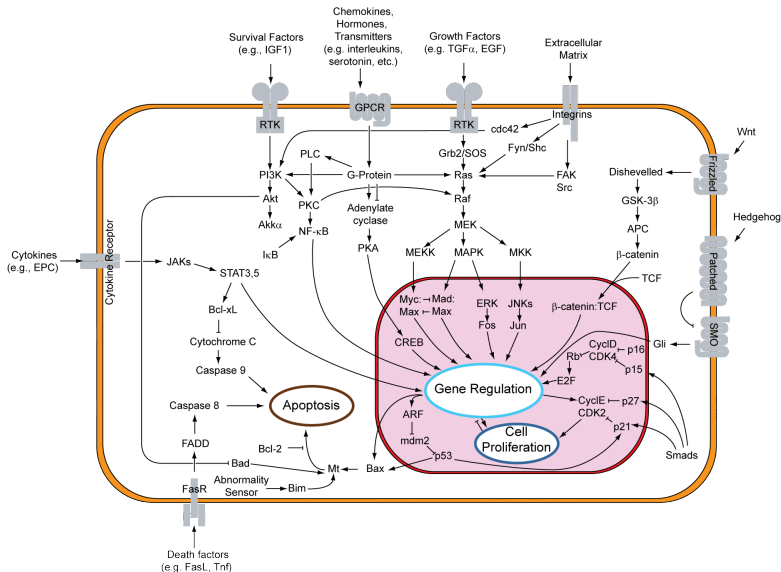
Врсте биолошких мрежа

- Биолошке мреже могу се поделити према величини и врсти основних градивних јединица моделованог биолошког система, представљених чвором графа, као и природи интеракције.
- Мале (микроскопске) мреже:
 - регулаторне мреже гена (*gene regulatory network, GRN*),
 - мреже преноса сигнала (*signal transduction network, STN*),
 - мреже протеинских интеракција (*protein interaction network, PIN*),
 - метаболичке мреже (*metabolic network*).
- Велике (макроскопске) мреже:
 - филогенетске мреже (*phylogenetic network*),
 - еколошке мреже (*ecological network*).
- Постоје и разне друге врсте мрежа, нпр. мождана (конектом).

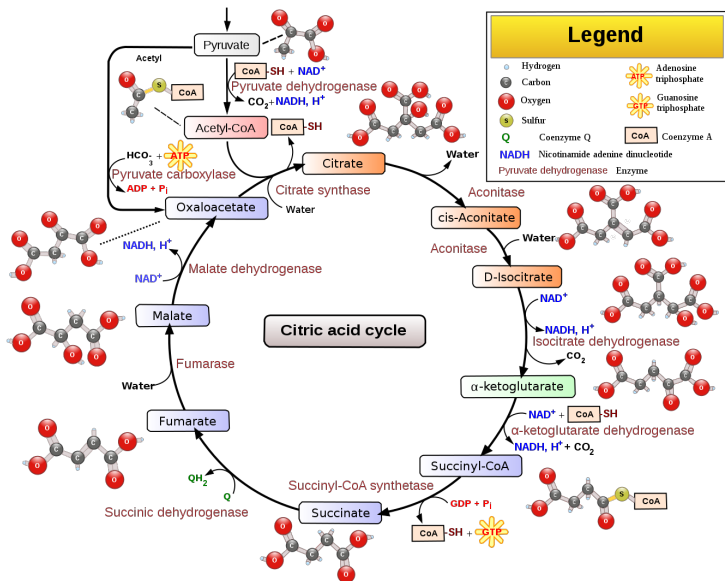
Пример генске регулације



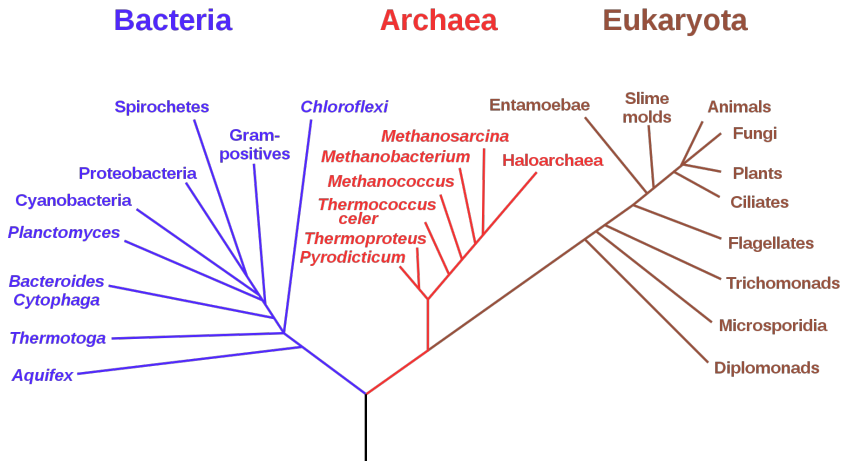
Сигнални путеви људске ћелије



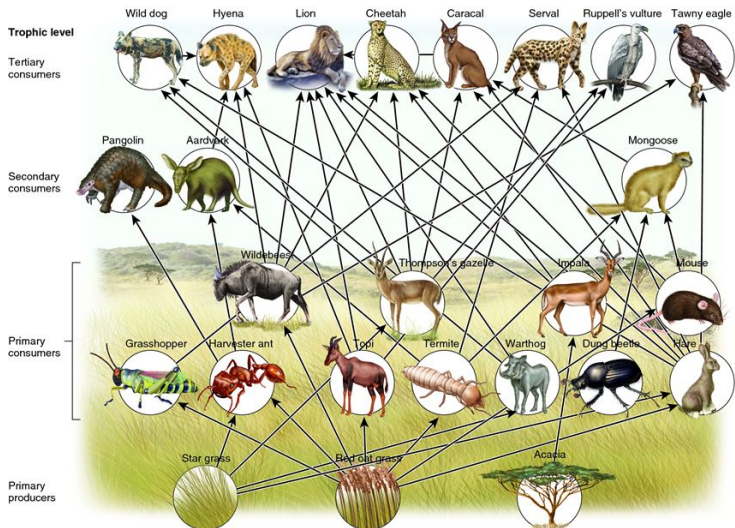
Кребсов циклус



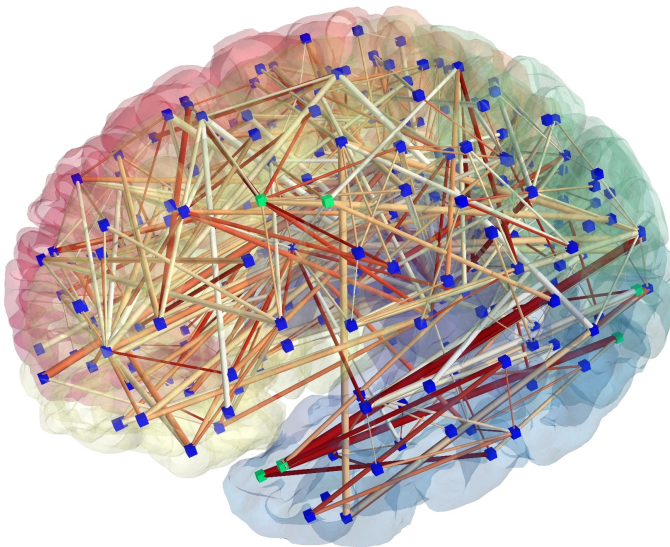
Дрво живота



Мрежа исхране у савани



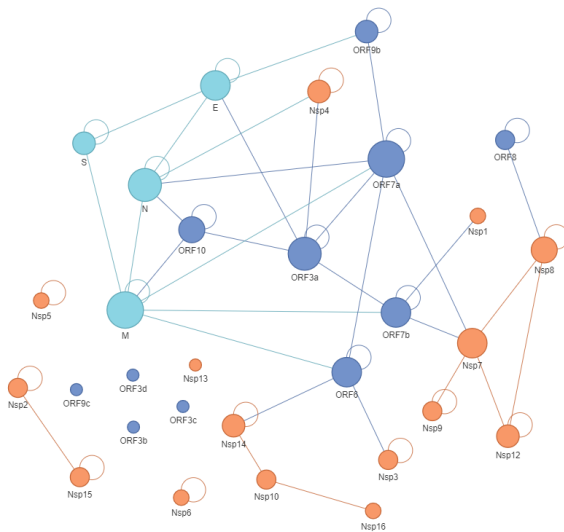
Људски конектом



Мреже протеинских интеракција

- У мрежама протеинских интеракција (*protein[-protein] interaction network*, $P[P]IN$), чворови су протеини, а гране неусмерене протеинске интеракције (*protein-protein interaction*, PPI).
- У PIN у ширем смислу, неки чворови могу бити молекули другог типа, а гране усмерене. У наставку се разматрају само уже $PPIN$ тј. PPI мреже, строго са протеинима и неусмереним везама.
- Мрежа протеинских интеракција се једном речју назива интерактом. Слично, њени чворови се називају интеракторима, а гране интеракцијама. Област проучавања је интерактомика.

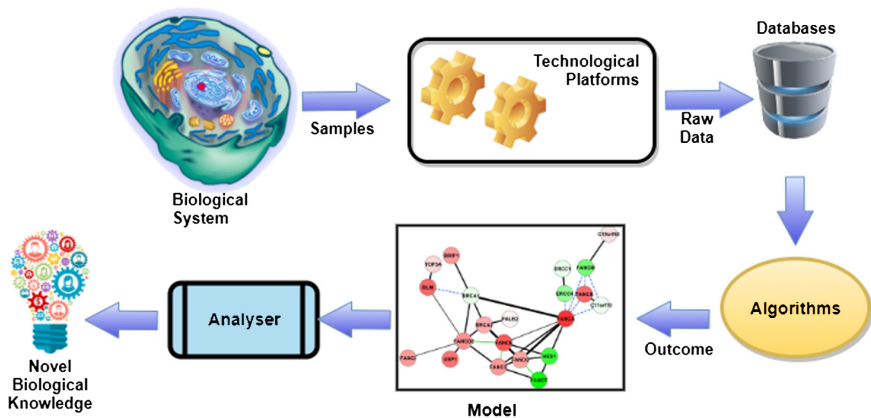
Интерактом SARS-CoV-2



Значај протеинских мрежа

- Протеини обављају скоро све процесе у организму и управљају њима, при чему врло ретко делују независно. Они од виталног значаја најчешће имају више интеракција (хабови), па се лако могу идентификовати увидом у одговарајућу мрежу интеракција.
- Поред тога, мреже се могу поредити („поравнати”) неким од алгоритама за утврђивање сличности графова, чиме се откривају нпр. хомологне подструктуре које су очуване током еволуције.
- Протеини који учествују у истим ћелијским процесима или су физички спојени често су густо повезани, па се тако из мреже могу издвојити протеински комплекси. Мана: модел бинарних интеракција је превише једноставан, па су многе везе сувишне. Решење: употреба степених графова (*power graph analysis*).

Ток података у интерактомици

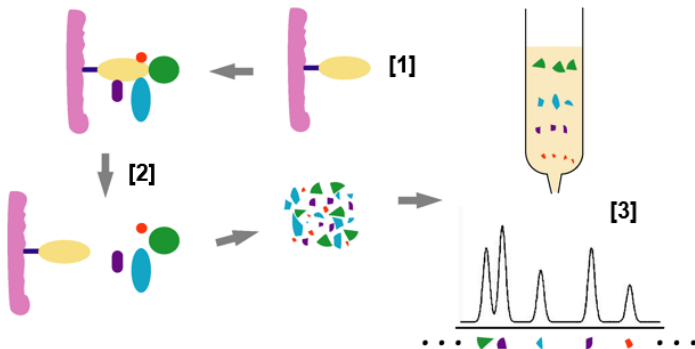


Откривање протеинских интеракција

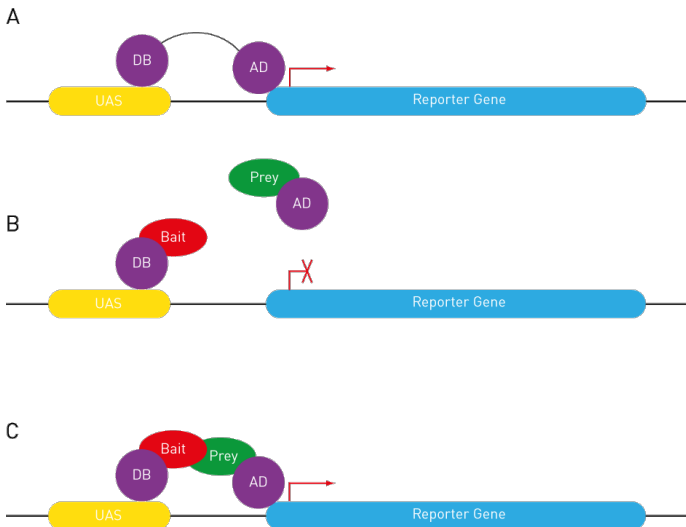
- Физичке интеракције између протеина махом се откривају експериментално, у лабораторији или пак рачунарски.
- Рачунарским (*in silico*) приступом могу се предвидети нове интеракције на основу старих, без потребе за лабораторијом. То нпр. може бити на основу сличности разматраних система.
- Лабораторијски приступ одликује се применом модела мамца и плена. Два најзаступљенија начина за одређивање да ли плен интерагује са мамцем јесу афинитетно прочишћавање праћено масеном спектрометријом (*affinity purification – mass spectrometry, AP-MS*) и двохибридна провера (*yeast-two-hybrid, Y2H*).

Прочишћавање и спектрометрија

Affinity purification+ mass spectrometry (AP-MS)



Двохибридна провера



Базе протеинских интеракција

- Интерактоми се могу генерисати упитом ка некој од база протеинских интеракција. Поред јавног интерфејса за дохватање списка интеракција, многе базе подржавају и графички приказ.
- Подаци о интеракцијама у базу се уносе ручно (експертска провера) или аутоматски, нпр. комбиновањем података из других база или анализом литературе методама истраживања текста.
- Неке базе (нпр. *IntAct*) интерактоме сматрају мултиграфима, при чему поновљене гране представљају везу која је потврђена у више извора. Неке базе (нпр. *STRING*) дају тежине гранама, које означавају степен сигурности у постојање интеракције.
- Још неке базе (*PSICQUIC*): *iRefIndex*, *MINT*, *BioGRID*...

PSICQUIC View

EMBL-EBI

Services
Research
Training
About us

PSICQUIC View

Examples: BRCA2 Q06609 dmc1,10831611

Input Form
Browse
Help

Feedback

12,808,706 binary interactions found for search term

☐ APID Interactomes

☒ BindingDB

☐ DIP-IMEx

☐ GeneMANIA

☐ InnateDB

☒ iRefIndex

☒ MINT

☐ Spike

☒ VirHostNet

☒ BAR

☒ BioGrid

☐ DrugBank

☒ HPIIDb

☐ InnateDB-All

☒ MatrixDB

☒ MPIDB

☒ tfact2gene

☐ ZINC

☒ bhf-ucf

☒ ChEMBL

☒ EBI-GOA-miRNA

☐ I2D

☒ IntAct

☒ MInfo

☒ Reactome

☐ TopFind

☐ BIND

☐ DiP

☒ EBI-GOA-nonIntAct

☒ IMEx

☐ Interoport

☒ mentha

☒ Reactome-Fls

☒ UniProt

Status of the service

☒ ONLINE

☐ OFFLINE

☐ WARNING: Time out

☐ ERROR: Unexpected Error

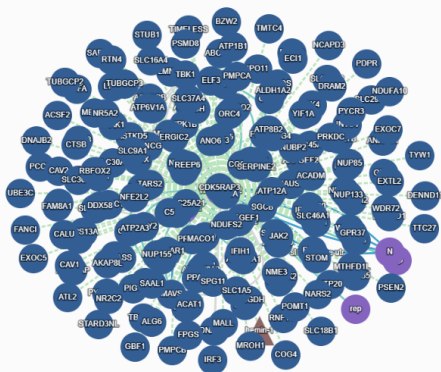
12,808,706 selected interactions

To many interactions to cluster. Please reduce the number to less than 5000 interactions.

version: 1.4.11

Interaction Network

Interactor Name



Неуређеност у интерактому

- Неуређени протеини учествују у многобројним ћелијским процесима, па су често део великог броја интеракција. Ово значи да би потенцијално могли имати велики степен у интерактому.
- Проучени су интерактоми генерисани на основу одабраних протеина *SARS-CoV-2*, а у њима су као особине од интереса издвојени степени повезаности и неуређености протеина.
- Полазни протеини: мембрански (M) и неструктурни (Nsp). Базе података: *IntAct* и *iRefIndex*. Разматране мере неуређености: аминокиселински профил, *IUPred*, *PONDR* и многе друге.

Jupyter свеска са резултатима

Neuređenost proteina u interaktomu

U radu je razmotren pojam interaktoma kao mreže proteinskih interakcija, kao i osobine i odnosi proteina u njemu. Konkretno, proučeni su interaktomi generisani na osnovu odabranog podskupa proteoma virusa SARS-CoV-2 (membranski i nestrukturirani proteini), a kao osobina od interesa izdvojena je neuređenost proteina u njima (prema različitim kriterijumima). Zadatak je bio uporediti stepen neuređenosti sa stepenom povezanosti čvora u grafu.





Sadržaj

1. [Proteinska sekvenca](#)
2. [Neuređenost prema IUPred](#)
3. [Interaktom prema IntAct](#)
4. [Poređenje osobina](#)
5. [Interaktom prema RefIndex](#)
6. [Virusni interaktom](#)
7. [Neuređenost prema profilu](#)
8. [Neuređenost prema POND](#)
9. [Dužina neuređenih regiona](#)
10. [Drugi kraj interakcije](#)
11. [Dužina sekvence](#)
12. [Najpovezaniji proteini](#)

Закључак

- Будући интерагујући блокови, биолошки системи се природно моделују графовима (мрежама). Резултујуће мреже су разноврсне по природи, али имају специфичне тополошке карактеристике.
- Мреже протеинских интеракција (интерактоми) моделују односе протеина (интерактора). Њиховом биоинформатичком анализом могуће је доћи до нових знања о разматраним протеинима.
- На примеру анализе неуређености протеина у интерактомима генерисаним на основу одабраних протеина вируса *SARS-CoV-2*, приказан је целокупан ток података у интерактомици.

Литература

-  Dmitrij Frishman and Manja Marz, *Virus bioinformatics*, CRC Press, 2021.
-  Pietro Hiram Guzzi and Swarup Roy, *Biological network analysis*, Elsevier Inc., 2020.
-  Björn H. Junker and Falk Schreiber, *Analysis of biological networks*, John Wiley & Sons, Inc., 2008.
-  Birthe B. Kragelund and Karen Skriver, *Intrinsically disordered proteins*, Springer Science+Business Media, LLC, 2020.