

# Sadržaj

<b>1. Uvod: objašnjavanje bioloških mreža teorijom grafova</b>	<b>6</b>
1.1. Uvod . . . . .	6
1.2. Poređenje mreža preko njihovih svojstava . . . . .	9
1.3. Modeli mreža . . . . .	12
1.4. Poravnavanje mreža . . . . .	14
1.5. Topologija mreže i biološke funkcije i bolesti . . . . .	17
1.6. Prognoza . . . . .	19
<b>2. Modelovanje interaktoma: bez skale ili geometrijsko?</b>	<b>21</b>
2.1. Uvod . . . . .	21
2.2. Sistem i metode . . . . .	24
2.2.1. Analiza graflata PPI mreža . . . . .	24
2.3. Rezultati i diskusija . . . . .	26
2.3.1. Utvrđivanje frekvencije graflata . . . . .	26
2.3.2. Robusnost utvrđivanja frekvencije graflata . . . . .	27
2.3.3. Globalna svojstva PPI mreža i njihovih modela . . . . .	27
2.4. Zaključci . . . . .	29
2.5. Zahvalnice . . . . .	31
<b>3. Raspodela stepena graflata</b>	<b>32</b>
3.1. Uvod . . . . .	32
3.1.1. Istorijat . . . . .	33
3.2. Pristup . . . . .	35
3.2.1. PPI i modeli mreža . . . . .	35
3.2.2. Raspodela stepena graflata (GDD) . . . . .	36
3.2.3. „GDD slaganje“ mreža . . . . .	37
3.3. Rezultati i diskusija . . . . .	41
3.4. Zaključak . . . . .	43
<b>4. Potpisi stepena graflata i biološka funkcija</b>	<b>44</b>
4.1. Uvod . . . . .	44
4.1.1. Istorijat . . . . .	44
4.1.2. Pristup . . . . .	45

4.2. Metode . . . . .	46
4.3. Rezultati i diskusija . . . . .	48
4.3.1. Rezultati . . . . .	48
4.3.2. Diskusija . . . . .	56
4.3.3. Smernice za dalji rad . . . . .	56
4.4. Zaključci . . . . .	59
<b>5. Fitovanje proteinskih mreža geometrijskim grafovima</b>	<b>61</b>
5.1. Uvod . . . . .	61
5.2. Teorijska osnova . . . . .	62
5.3. Početni algoritam . . . . .	64
5.4. Algoritam za praktičnu primenu . . . . .	66
5.5. Podaci i rezultati . . . . .	69
5.5.1. PPI mreže . . . . .	69
5.5.2. Rezultati . . . . .	70
5.5.3. Dalja ispitivanja . . . . .	72
5.6. Diskusija . . . . .	74
<b>6. Geometrijsko prečišćavanje mreža proteinskih interakcija</b>	<b>77</b>
6.1. Uvod . . . . .	77
6.1.1. Mreže proteinskih interakcija . . . . .	77
6.1.2. Model geometrijskog grafa . . . . .	79
6.2. Metode . . . . .	80
6.2.1. Algoritam ugrađivanja . . . . .	81
6.2.2. Geometrijsko uklanjanje šuma u PPI mrežama . . . . .	82
6.2.3. Podaci . . . . .	85
6.3. Rezultati . . . . .	85
6.3.1. Ispitivanje geometrijskog prečišćavanja/uklanjanja šuma . . . . .	85
6.3.2. Primena na ljudsku PPI mrežu . . . . .	87
6.4. Diskusija . . . . .	90
6.5. Slike . . . . .	92
<b>7. Topološko poravnavanje mreža otkriva biološku funkciju i filogeniju</b>	<b>96</b>
7.1. Uvod i motivacija . . . . .	96
7.1.1. Teorijska osnova . . . . .	97
7.1.2. Prethodni pristupi . . . . .	97
7.1.3. Naš doprinos . . . . .	99
7.2. Rezultati i diskusija . . . . .	100
7.2.1. Procena algoritma GRAAL . . . . .	101
7.2.2. Poravnavanje PPI mreža kvasca i čoveka po parovima . . . . .	102

---

7.2.3.	Kvalitet poravnavanja mreža kvasca i čoveka pomoću algoritma GRAAL . . . . .	103
7.2.4.	Poređenje s drugim metodama . . . . .	106
7.2.5.	Primena na predviđanje funkcije proteina . . . . .	107
7.2.6.	Rekonstrukcija filogenetskih stabala uz korišćenje poravnavanja metaboličkih putanja između bioloških vrsta . . . . .	109
7.3.	Zaključci . . . . .	111
7.4.	Metode . . . . .	111
7.4.1.	Potpisi stepena grafleta i sličnosti potpisa . . . . .	112
7.4.2.	Algoritam GRAAL (GRAph ALigner) . . . . .	113
7.4.3.	Statistički značaj našeg poravnavanja mreža kvasca i čoveka . . . . .	114
<b>8.</b>	<b>Integrativan pristup modelovanju bioloških mreža</b> . . . . .	<b>119</b>
8.1.	Uvod . . . . .	120
8.1.1.	Istorijat . . . . .	120
8.1.2.	Naš doprinos . . . . .	122
8.2.	Metode . . . . .	123
8.2.1.	Skupovi podataka . . . . .	123
8.2.2.	Tehnike . . . . .	124
8.3.	Rezultati i diskusija . . . . .	126
8.3.1.	Rezultati . . . . .	126
8.3.2.	<b>Robustnost i validacija</b> . . . . .	<b>128</b>
8.3.3.	Poređenje s drugim istraživanjima . . . . .	131
8.3.4.	Diskusija . . . . .	132

# Predgovor

Proteini su trodimenzioni makromolekuli koji obavljaju sve ćelijske funkcije vezujući se fizički jedan za drugi. Skup svih proteina u ćeliji i svih interakcija između njih čini mrežu proteinskih interakcija. Ove mreže su ključne za razumevanje svih bioloških procesa u ćeliji. Pošto lekovi deluju na proteine, razumevanje ovih mreža će omogućiti sintezu novih lekova za hemijsku manipulaciju ovih mreža koja doprinosi izlečenju. Trenutno su javno dostupni eksperimentalni podaci o ovim mrežama za više organizama. Mnoge laboratorije danonoćno rade na proizvodnji novih podataka i kompletiranju ovih mreža.

Na žalost, mnogi problemi koji se bave mrežama su računski neizvodljivi (NP-teški ili NP-kompletni). Ovo znači da je praktično nemoguće tačno rešiti te probleme. Pošto su ove mreže ključ za razumevanje života, privrgavamo dizajniranju metoda za približno rešavanje ovih problema (tzv. heuristike). Međutim, pošto ne postoji heuristika koja dobro aproksimira rešenja u svakom mogućem slučaju, modeliranje podataka je bitno da bi se dizajnirale heuristike za koje možemo dokazati da dobro aproksimiraju rešenje na datom modelu.

U ovoj monografiji, dajem kratak pregled svog naučnog rada od 2004. do 2010. godine na modeliranju i algoritamskom razvoju metoda za analiziranje mreža proteinskih interakcija. Počevši sa uvodom (glava 1), u kome dajem kratak pregled glavnih problema i trenutno dostupnih metoda za njihovo rešavanje, kao i nekih polemika u polju, monografija nastavlja sa opisom novih metoda za analiziranje i modeliranje ovih mreža (glave 2, 3, 4 i 5). U šestoj glavi opisujemo primenu modela za uklanjanje šuma iz ovih mreža. Sedma glava predstavlja algoritam za poravnavanje proteinskih mreža raznih vrsta i njegovu primenu za predviđanje funkcije trenutno nekarakterizovanih proteina, kao i za rekonstrukciju filogenije pomoću poravnavanja mreža umesto pomoću poravnavanja genetičkih sekvenci koje je trenutni standard u disciplini. Na kraju, predstavljamo rad na integrativnom modeliranju ovih mreža.

Zahvaljujem akademiku Prof. Dragošu Cvetkoviću na korisnim komentara-

rima koji su mi pomogli u pisanju ove knjige. Takođe zahvaljujem Prof. Stevanu Milinkoviću na pomoći oko objavljivanja, kao i Olgi Milanko na prevodima tekstova sa engleskog. Zahvaljujem Dr. Srđanu Marinoviću na tehničkoj pomoci oko L<sup>A</sup>T<sub>E</sub>X-a kao i na prijateljskoj podršci. Nadam se da će ova monografija približiti čitaocu novu oblast istraživanja mreža proteinskih interakcija.

London, 11. septembar 2011. godine.

Autorka

# Glava 1. Uvod: objašnjavanje bioloških mreža teorijom grafova

Nova, sve razvijenija oblast bioloških mreža trebalo bi da pruži uvid u principe organizacije života. Međutim, uprkos značajnim zajedničkim naporima naučnika, biolozi i informatičari su i dalje slabo povezani, i nema multidisciplinarnog razumevanja predmeta istraživanja. Posledica toga je korišćenje jednostavnih računarskih tehnika ograničenih mogućnosti, pomoću kojih je nemoguće objasniti te kompleksne podatke. Stoga postoji opasnost da će zajednica početi da posmatra topološke osobine podataka dobijenih iz mreža samo kao statistike, umesto kao bogate izvore bioloških informacija. Druga opasnost je da takav način posmatranja dovede do nametanja naučnih doktrina u ovoj oblasti, kao što su mišljenja kojima se na strani modelovanja potenciraju mreže „bez skale“ (engl. “scale-free” networks), a na strani biologije ona kojima se potencira genom kao jedini izvor bioloske informacije. U ovom poglavlju, mreže proteinskih interakcija razmotrene su sa aspekta teorije grafova, i dat je opšti pregled ove naučne oblasti, uz osvrt na potencijalne buduće izazove. Naše metode navedene u ovom poglavlju su detaljno opisane u narednim poglavljima.

## 1.1. Uvod

Zahvaljujući nedavnim inovacijama u eksperimentalnoj biologiji, pojavile su se velike količine podataka o interakcijama između gena i proteina. Brzim automatizovanim biohemijskim analizama, kao što su analize dvohibridnog kvasca (engl. yeast-two-hybrid, Y2H) [92, 207, 128, 71, 190, 171, 183], afinite-tna hromatografija s masenom spektroskopijom (engl. affinity purification coupled to mass spectrometry, AP/MS [69, 88, 70, 117] i mreže sintetičkih letalnih i supresorskih interakcija [206], dobijene su parcijalne mreže za humane [190, 171], mikrobne [165, 151, 124] i virusne [26] patogene, kao i za mnoge model organizme [92, 207, 71, 128, 70, 117, 206]. Ove velike mreže pružaju mnoge zanimljive i važne mogućnosti biologima i informatičarima.

Živimo u jedinstvenom vremenu u istoriji nauke, kada dostignuća u razvoju algoritama i modelovanja primenjena na ove podatke mogu doprineti boljem razumevanju biologije i potencijalno uticati na metode lečenja i javno zdravlje.

Pred novonastajućom naučnom oblašću bioloških mreža stoje značajni izazovi. Prvo, naši tekući opservacioni podaci u velikoj meri su nepotpuni zbog načina uzorkovanja, uprosečavanja populacije i drugih odstupanja u prikupljanju, obradi i tumačenju podataka [213, 43, 193, 80, 44, 36, 78, 223]. Uz to, u njima ima „šuma“ (tj. nerelevantnih, neobjašnjivih i/ili nestrukturiranih podataka) zbog ograničenja biotehnoških metoda korišćenih za njihovo prikupljanje. Primer mreže proteinskih interakcija (engl. protein interaction network, PIN; ili protein-protein interaction network, PPI network) koji ilustruje oskudnost podataka prikazan je na slici 1. Uprkos nekompletnosti i šumu, naučna zajednica je počela da analizira i modeluje te podatke, što je dovelo do zanimljivih i ponekad kontroverznih otkrića [97, 199, 96, 40, 188, 93, 108, 55, 60, 58, 59]. Na primer, postavlja se pitanje da li su metaboličke mreže „bez skale“ ili imaju „gustu skalu“ (engl. “scale-rich”) [97, 199], da li su najpovezaniji čvorovi u mreži proteinskih interakcija kvasca *Saccharomyces cerevisiae* letalni [96, 40], da li su složene mreže „sebi slične“ (engl. “self-similar”) ili „nisu sebi slične“ (engl. “self-dissimilar”) [188, 93], te da li se čvorovi u mreži proteinskih interakcija kvasca mogu podeliti na „raskalašne“ čvorove (engl. “party hubs”), one koji simultano stupaju u interakcije sa svojim partnerima, i „pristojne“ čvorove – (engl. “date hubs”), tj. one koji stupaju u interakcije s različitim partnerima u različito vreme ili na različitim mestima [55, 60, 58, 59] (videti odeljak 1.5.). Kontroverze su često posledica slabog razumevanja svojstava podataka koji su posledica uzorkovanja podataka, kao i upotrebe računarskih tehnika osetljivih na šum [80, 193, 44]. Uz to, na početku je veliki značaj bio pripisivan raspodelama stepena cvorova po stepenoj funkciji (engl. power law of the degree distribution) u mnogim biološkim mrežama, npr. sugerišući „univerzalnu arhitekturu“ složenih sistema, što je kasnije osporeno [108].

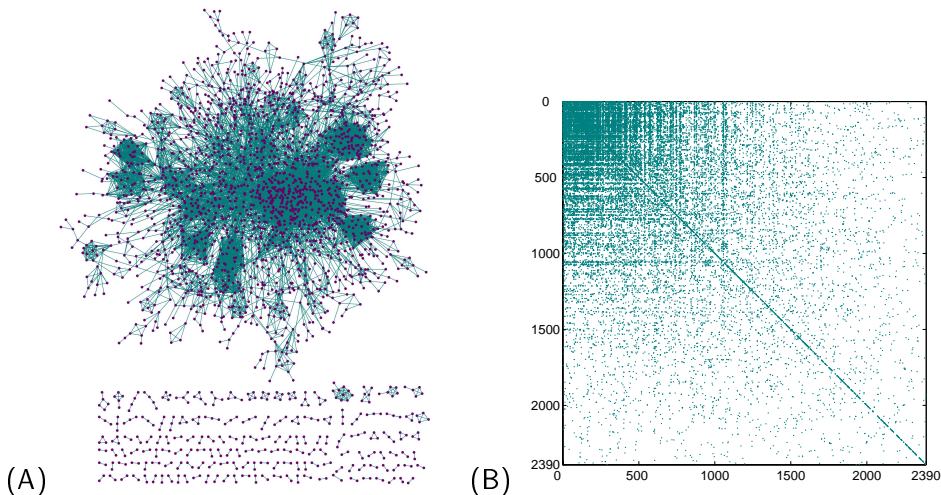
Modelovanje bioloških mreža je važno, pošto samo pomoću modela možemo dobiti sažet pregled fenomena u istraživanju koje može pokazati neke neočekivane osobine koje bismo možda želeli da dalje eksperimentalno proučavamo. Postoje različite paradigme modelovanja. U ovom poglavlju usredsređujemo se na analizu i modelovanje pomoću teorije grafova. Može se tvrditi da su takve analize previše pojednostavljen pogled na mrežne sisteme pošto se radi o apstrakcijama u kojima se zanemaruje dinamika procesa (npr. videti [90, 77] s razmatranjem dinamike). Mnoge proteinske interakcije su prirodno neusmerene i predstavljaju takozvane „stabilne“ interakcije, u kojima proteini koji stupaju u interakcije pronalaze jedni druge i ostaju

povezani (npr. proteinski kompleksi kao što su ribozom ili hemoglobin), dok su druge „prolazne“ interakcije jer se proteini koji stupaju u interakcije povezuju i rastavljaju u različito vreme i pod različitim uslovima (npr. signalne kaskade). Stoga, kada bismo potpuno poznavali neku mrežu proteinskih interakcija, ona bi morala da bude predstavljena kao kombinovana mreža usmerenih i neusmerenih veza, s komponentama koje zavise od vremena a obuhvataju mnoge parametre. Međutim, takve informacije su trenutno nedostupne na nivou sistema, pa su mreže proteinskih interakcija statičke neusmerene mreže u kojima se kombinuju stabilne interakcije sa svim trenutno poznatim prelaznim interakcijama, u svim ispitivanim trenucima i eksperimentalnim uslovima. Ipak, čak i jednostavna analiza takvih podataka zasnovana na teoriji grafova već je pružila vredan uvid u biološke sisteme (videti niže). Stoga se može tvrditi da ti podaci sadrže značajne biološke informacije skrivene u njihovoj složenosti koje se mogu otkriti razvojem i primenom sofisticiranih tehnika teorije grafova.

Cilj ovog poglavlja nije da polemiše o redukcionističkom i determinističkom modelovanju (koje se trenutno u filozofiji nauke smatraju naivnim ili pojednostavljenim) nasuprot dinamičkom modelovanju, o pitanjima odnosa između teorije, modela i eksperimenta, niti o tome šta se tačno može „predvideti“ pomoću modela ili teorije; npr. videti [107, 48]. Umesto toga, tvrdim da je analiziranje i modelovanje *topologije* (ili i *strukture*) bioloških mreža jedan od načina na koji se može izvući biološka informacija iz mrežnih podataka. Shodno tome, neću prezati od korišćenja analogije sa fizikom i astronomijom u 17. veku: mada su, zahvaljujući Koperniku, Kepleru, Galileju i drugima, u to vreme bili na raspolaganju dobri opservacioni podaci zajedno sa suprostavljenim teorijama o tome da li je naš solarni sistem geocentričan ili heliocentričan, tek kada je ser Isak Njutn formulisao opšte zakone gravitacije i kretanja, razumeli smo zašto se planete kreću tako kako se kreću. U svetlu tih zakona opservacije o našem solarnom sistemu su postale očigledne i razumeli smo da je naš solarni sistem samo deo mnogo većeg fenomena koji se može opisati istim zakonima. Slično tome, glavni razlog modelovanja podataka pomoću mreža jeste da bismo razumeli zakone, pošto se samo uz korišćenje zakona možemo nadati da ćemo moći da predviđamo i reprodukujemo fenomene. Jedan od načina modelovanja je korišćenje teorije grafova. Uprkos nesavršenim podacima, svojstva mrežnog modela već su korišćena za smanjivanje složenosti. Na primer, korišćena su da bi se predložila strategija za optimalno otkrivanje interaktoma [125]. Osim toga, koristili smo ih i za razvijanje efikasnih algoritama za aproksimativno rešavanje problema koji se teško rešavaju matematički (takvi algoritmi se zovu *heuristički* algoritmi) [160] kao i za uklanjanje šuma iz podataka



[120].



Slika 1 **A.** Primer PIN mreže [36]. **B.** Matrica susedstva iste mreže koja ilustruje njenu oskudnost.

## 1.2. Poređenje mreža preko njihovih svojstava

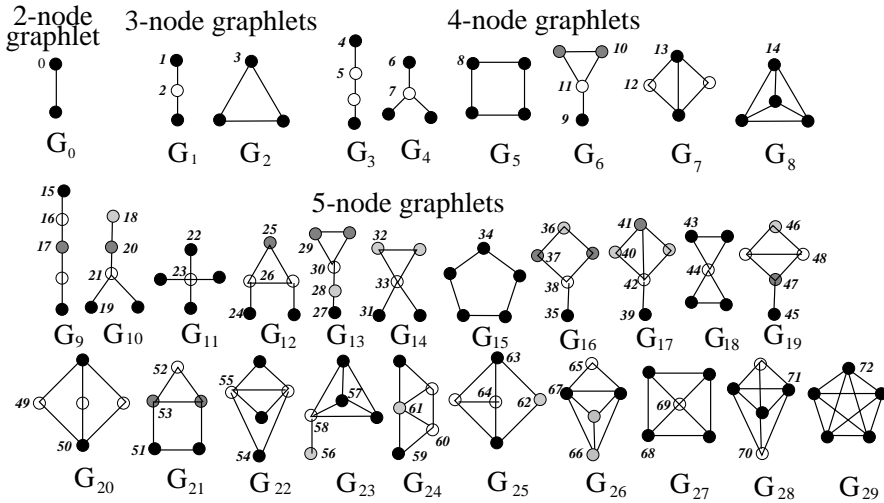
Usporedna analiza mrežnih podataka omogućila bi pronalaženje sličnosti i razlika između bioloških mreža, kao i prenos znanja. To bi bilo korisno, pošto se često dosta toga zna o mreži model organizma, ali vrlo malo o mrežama drugih organizama. Međutim, poređenje velikih mreža (koje se zovu i *grafovi*) nije računski izvodljivo, jer bi za takvo izračunavanje bilo potrebno rešiti „problem izomorfizma podgrafova“ (engl. “subgraph isomorphism problem”) u kome se postavlja pitanje da li jedan graf postoji kao kopija u drugom grafu; nažalost, taj problem je računarski teško rešiv (NP-kompletan), pa nema efikasnog načina za njegovo rešavanje. Zbog toga se traže lako izračunljivi heuristički pristupi.

Neki od ovih heurističkih pristupa zovu se *svojstva mreža* i mogu se grubo i hronološki podeliti u dve grupe: makroskopska statistička *globalna svojstva* mreža, i mikroskopska *lokalna svojstva* mreža. U najčešće korišćena globalna svojstva mreža spadaju *raspodela stepena*, *koeficijent klasterovanja*, *spektar klasterovanja*, *prečnik mreže* i različiti oblici *centraliteta* mreže [147]. Bazirano na ovim svojstvima, predloženi su modeli bioloških mreža (videti odeljak 1.3.) ako svojstva mreža koje pripadaju modelu odgovaraju svojstvi-

ma mreža podataka. *Stepen* čvora je broj grana koje čvor dodiruje; prema tome, stepen je *svojsstvo čvora*. *Raspodela stepena* je raspodela stepena svih čvorova u mreži; isto tako, to je verovatnoća da slučajno izabran mrežni čvor ima stepen  $k$  (što se obično označava sa  $P(k)$ ). Mnoge biološke mreže, uključujući PIN-ove, imaju raspodelu koja ne odgovara Poasonovoj raspodeli, nego im rep sledi stepenu funkciju,  $P(k) \sim k^{-\gamma}$ ,  $\gamma > 0$ ; sve takve mreže su nazvane *mrežama bez skale* [12]. Zahvaljujući konceptualnoj jednostavnosti ovog svojstva, ono je masovno korišćeno za „karakterizaciju“ stvarnih mreža, iako ono veoma slabo opisuje strukturu mreže [159, 127]. Na primer, mreža sačinjena od 5 trouglova i mreža koju čini jedan prsten sa 15 čvorova iste su veličine, tj. imaju isti broj čvorova i grana, kao i identicnu raspodelu stepena (svaki čvor je stepena 2), ali se njihove topologije suštinski razlikuju. Slično važi i za ostala globalna svojstva mreže, uključujući koeficijent klasterovanja, koji je mera „grupisanosti“ mreže, kao i za prečnik mreže, koji meri koliko se daleko „prostire“ mreža [159]. PIN-ovi i druge biološke i stvarne mreže imaju velike koeficijente klasterovanja u poređenju s potpuno slučajnim mrežama, i mali prosečan prečnik, reda veličine  $O(\log n)$ , gde je  $n$  broj čvorova u mreži; to svojstvo mreže zove se svojstvo *malog sveta* (engl. “small world”) [219]. Međutim, globalna svojstva u velikoj meri nekompletnih PIN-ova mogu proizvesti pogrešne interpretacije pošto opisuju strukturu nekompletnih mreža dobijenih uzorkovanjem, a ne pravu strukturu trenutno nepoznatih kompletnih mreža [80, 193, 44]. Srećom, neki lokalizovani delovi PIN-ova dobro su proučeni – to su obično delovi mreža značajne za bolesti ljudi. Prema tome, lokalna svojstva primenjena na dobro proučene lokalne oblasti mreže su mnogo prikladnija.

Lokalna svojstva koja razmatramo zasnivaju se na postojanju podgrafa. *Podgraf* (ili *parcijalni podgraf*, engl. “partial subgraph”) mreže  $G$  jeste mreža čiji čvorovi i grane pripadaju  $G$ . *Indukovani podgraf* (engl. “induced subgraph”)  $H$  mreže  $G$  mora da sadrži *sve* grane mreže  $G$  koje povezuju čvorove podgrafa  $H$ . Na primer, u grafu  $G_2$  na slici 2, ako izaberemo sva tri čvora, možemo odabrati bilo koje dve grane između njih da bismo formirali linearnu putanju sa 3 čvora (kao što je  $G_1$  na istoj slici); takve putanje sa 3 čvora parcijalni su podgrafovi grafa  $G_2$ . Pošto indukovani podgraf mora da sadrži sve grane, indukovani podgraf grafa  $G_2$  sa 3 čvora je trougao. Analogno motivima sekvenci, *motivi mreže* su definisani kao parcijalni podgrafovi koji se javljaju u stvarnim mrežama mnogo učestalije nego što bi se očekivalo u slučajnim mrežama (engl. “random graphs”) [143, 180, 142]. Predloženo je da su motivi mreža funkcionalni gradivni blokovi mreža regulacije gena i drugih bioloških mreža, kao i da su za različite tipove stvarnih mreža karakteristični različiti motivi [143, 180, 142, 5]. Međutim,

ovakva definicija motiva je nejasna, jer postoji mnogo različitih modela slučajnih grafova (videti odeljak 1.3.) [8]. Pored toga, za karakterizaciju strukture bilo koje familije mreža, koriste se indukovani, a ne parcijalni podgrafovi [22]. Zbog toga smo definišali *graflet* kao mali, povezan, **indukovan** podgraf mreže. Uveli smo analize zasnovane na učestalosti pojavljivanja *svih* grafleta koji imaju do 5 čvorova (predstavljeno na slici 2), bez obzira na to da li su oni previše zastupljeni u podacima u poređenju s bilo kojim modelom mreže [159, 158]. Znači, grafleti ne moraju da budu previše zastupljeni, i po tome se, zajedno s tim što su indukovani, razlikuju od motiva mreža.

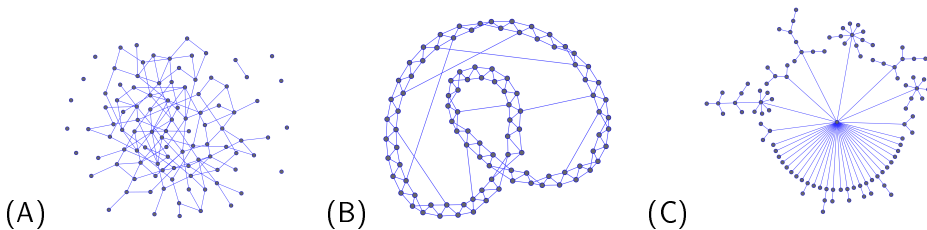


Slika 2 Svi grafleti sa 2, 3, 4 i 5 čvorova,  $G_0, G_1, \dots, G_{29}$ , i orbitama označenim sa  $0, 1, 2, \dots, 72$  [158]. Na svakom grafletu, čvorovi koji pripadaju istoj orbiti predstavljeni su istom nijansom.

Odredjivanjem frekvencije (učestalosti) grafleta u mreži, dobijamo statističke karakteristike lokalne strukture mreže nezavisno od bilo kog multog modela mreže, a poređenjem ovih raspodela frekvencije dobijamo meru sličnosti strukture mreža [159]. Metodama baziranim na grafletima smo dalje povećali senzitivnost time što smo zapazili „simetrije“ između čvorova u grafletima. Na primer, u prstenu (krugu) od četiri čvora ( $G_5$  na slici 2), svaki čvor izgleda isto kao svaki drugi, ali u lancu (na putanji) sa četiri čvora, postoje dva krajnja i dva središnja čvora. Tu ideju formalizujemo tako što koristimo „orbite automorfizma“ grafa (opisane niže) [158]. Na taj način povećavamo osetljivost upotrebe grafleta za analizu i modelovanje mreža, a da pri tom izračunavanje ne postaje skuplje.

### 1.3. Modeli mreža

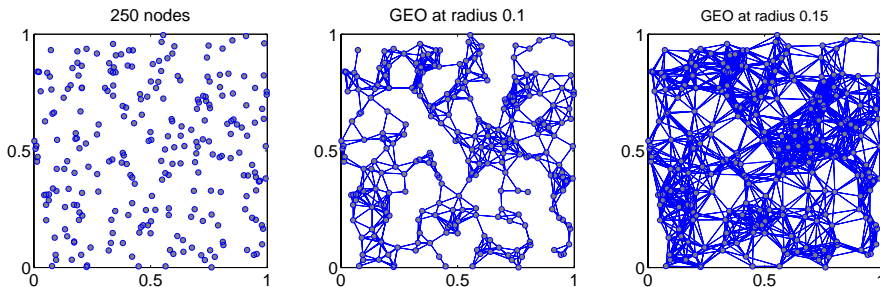
*Erdos-Renyi* („ER“) slučajni grafovi najstariji su model slučajnih grafova. U tim grafovima, grane između parova čvorova definisane su uniformno nasumično sa istom verovatnoćom,  $p$  [50]. Taj model je intenzivno proučavan i mnoga njegova svojstva su matematički opisana [20]. Zbog toga je to standardni model s kojim se porede podaci, bez obzira na to što se ne očekuje da se ovaj model dobro slaže s podacima. Pošto ER grafovi, za razliku od PIN-ova, imaju Poasonovu raspodelu i male koeficijente klasterovanja, definisani su drugi modeli mreža od kojih se očekivalo da će dobro fitovati podatke. U *uopštenim slučajnim grafovima* (engl. “generalized random graphs”), grane su izabrane nasumično kao i u ER grafovima, ali je raspodela stepena uslovljena da odgovara raspodeli stepena podataka [148, 3]. Mreže *malog sveta* integrišu pravilnost i nasumičnost jer su to pravilni prsteni sa rešetkom, s malim brojem nasumično prepovezanih grana; stoga one imaju male prečnike i velike koeficijente klasterovanja [219]. Mreže *bez skale* („SF“) sadrže i dodatno ograničenje: raspodela stepena prati stepenu funkciju [12, 127]. Pošto raspodele stepena mnogih PIN-ova opadaju približno po stepenoj funkciji, predložene su mnoge varijante metoda za generisanje SF mreža, od kojih su za PIN-ove najznačajnije one zasnovane na biološki motivisanim principima duplikacija i mutacija gena kroz evoluciju [215, 152]. Primeri modela mreža dati su na slici 3.



Slika 3 Primeri modela mreža. (A) Erdos-Renyi slučajni graf. (B) Mreža malog sveta. (C) Mreža bez skale.

Predložili smo alternativne modele: biološki motivisane modele PIN-ova zasnovane na *geometrijskim grafovima* („GEO“) [159, 158, 119, 162, 133]. Pretpostavimo da imamo skup tačaka raspoređenih u prostoru. Biramo konstantno rastojanje  $\epsilon$  i kažemo da su dve tačke „povezane“ ukoliko su unutar rastojanja  $\epsilon$  jedna od druge. Taj odnos se može predstaviti grafom u kome je svaka tačka u prostoru čvor, a dva čvora su povezana granom ako su im odgovarajuće tačke u prostoru na rastojanju manjem ili jednakom  $\epsilon$  jedna

od druge. Takav graf se zove *geometrijski graf*; ukoliko su tačke raspoređene nasumično u prostoru, onda je to *geometrijski slučajni graf* (engl. “geometric random graph”). Ilustracije geometrijskih grafova su prikazane na slici 4.



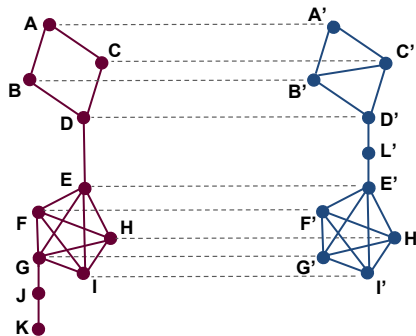
Slika 4 Levo: jedinična kocka ima 250 tačaka. U sredini: rezultujući geometrijski graf sa graničnim rastojanjem od 0.1. Desno: iste tačke ali drugačiji graf, sa graničnim rastojanjam od 0.15.

Kao što je gore pomenuto, raspodela stepena veoma slabo karakteriše strukturu grafa. Metode zasnovane na grafitima postavljaju znatno strožije zahteve. Koristeći te metode, pokazali smo da GEO modeli mnogo bolje fituju podacima nego SF modeli [159, 158]. Zbog toga se pojavljuje pitanje kako odlučiti kom svojstvu verovati pri poređenju mreža. Pokazali smo sledeće: kada se niz globalnih i lokalnih svojstava mreža koristi za poređenje raznih PIN-ova sa različitim modelima mreža, primenom nekoliko klasifikatora za mašinsko učenje (engl. “machine learning classifiers”) na ta svojstva potvrdili smo da je struktura PIN-ova najsličnija strukturi GEO mreža sa šumom [133]. To smo dodatno verifikovali tako što smo pokazali da se PIN-ovi mogu eksplicitno ugraditi u niskodimenzioni geometrijski prostor [86]; da bismo izgradili GEO mrežu koja odgovara PIN-u, kao rastojanje između proteina koristili smo funkciju najkraće putanje između njih u datom PIN-u. Dodatno smo poboljšali GEO model kako bi još bolje odgovarao PIN-ovima na dva načina: prvo, tako što smo naučili kakva je raspodela proteina ugrađenih u prostor pomoću ove metode [119] i koristili tu raspodelu za generisanje novih GEO mreža, i drugo, tako što smo u geometrijskom prostoru simulirali gorepomenute evolucionarne principe duplikacije i mutiranja gena [162]. Konceptualno, dobro slaganje GEO mreže s PIN-ovima moglo bi se objasniti zapažanjem da se svi biološki entiteti, uključujući gene i proteine kao produkte gena, nalaze u nekom višedimenzionalnom biohemijskom prostoru; u ovom trenutku je teško pretpostaviti prirodu tog prostora. Duplikacije i mutacije gena prirodno se mogu modelovati u tom biohemijskom prostoru:

duplicirani gen egzistira u istoj tački u prostoru kao njegov roditelj, a zatim se prirodnom selekcijom ili eliminiše jedan, ili se oni polako udaljavaju u prostoru jedan od drugog. To znači da gen „dete“ nasleđuje interakcije s nekim proteinima od gena „roditelj“, ali da može i da uspostavi veze sa drugim proteinima. Što se dete više udalji od roditelja u ovom apstraktnom prostoru, to će njihova biohemijska svojstva biti različitija. U ovom trenutku, takvi GEO modeli su prilično grube matematičke aproksimacije stvarne biologije, pa ih treba dalje poboljšavati.

## 1.4. Poravnavanje mreža

Mreže se mogu porediti i „poravnavanjem“ (engl. “network alignment”). Analogno poređenju i poravnavanju sekvenci (engl. “sequence alignment”), koje je značajno uticalo na naše razumevanje evolucije, biologije i bolesti, poređenje i poravnavanje bioloških mreža verovatno će imati sličan uticaj. Stoga se smatra da je pronalaženje metoda za pouzdano poređenje mreža različitih organizama jedan od najvažnijih problema evolucione biologije i biologije sistema (engl. “systems biology”) [178]. Konceptualno, poravnavanjem mreža se pokušava pronaći najbolji način za „uklapanje“ mreže  $G$  u mrežu  $H$ , čak i ako  $G$  ne postoji kao stvarni podgraf mreže  $H$  [178]. Jednostavan primer koji ilustruje poravnavanje mreža prikazan je na slici 5. Nažalost, za razliku od poravnavanja sekvenci, svaku razumnu formulaciju ovog problema nemoguće je tačno rešiti. Razlog za to je goreopisan problem izomorfizma podgrafova. Uz to, PIN-ovi i druge biološke mreže sadrže šum, tj. mnoge grane nedostaju, a mnoge su lažno prisutne [212]. Osim toga, biološke varijacije dodatno otežavaju merenje „uspešnosti“ poravnavanja mreža.



Slika 5 Primer poravnavanja dve mreže.

Kao i u slučaju poravnavanja sekvenci, postoje *lokalna* i *globalna* poravnavanja mreža. U lokalnim poravnavanjima, jedan čvor se može mapirati (preslikati) na nekoliko čvorova. Za razliku od toga, globalno poravnavanje mreža obezbeđuje jedinstveno poravnavanje: od svakog čvora u manjoj mreži do samo jednog čvora u većoj mreži. Međutim, mana je što to može dovesti do neoptimalnog preslikavanja čvorova u nekim lokalnim oblastima mreža. Većina metoda koje se koriste za poravnavanje usredsređena je na lokalna poravnavanja čiji je cilj da se pronađu male podmreže koje odgovaraju „bioloskim putanjama“ kroz ćeliju (engl. “biological pathways”), ili proteinskim kompleksima, s ocekivanjem da su očuvani u PPI mrežama različitih bioloških vrsta kroz evoluciju [109, 16, 64, 17]. Najstariji takav algoritam je PathBLAST, koji traži dobra poravnanja bioloških putanja između dva PIN-a tako što uzima u obzir verovatnoće da su proteinske interakcije na nekoj biološkoj putanji zaista prave interakcije a ne lažno pozitivni rezultati, kao i informacije o homologiji izvedene iz sekvenci poravnatih proteina [109]. Pomoću algoritma PathBLAST, identifikovane su ortologne biološke putanje između kvasca *S. cerevisiae* i bakterije *H. pylori*. Pomoću modifikacije algoritma PathBLAST, zvane NetworkBLAST-M, identifikovani su evolucionjski očuvani proteinski kompleksi kod više bioloških vrsta [176]. U drugom pristupu, zvanom MaWISH (Maximum Weight Induced Subgraph – indukovani podgraf najveće težine), koriste se modeli duplikacije i divergencije kako bi se razumela evolucija proteinskih interakcija [115]. Primenom ovog modela, MaWISH konstruiše graf globalnog poravnanja sa težinama (engl. “weighted global alignment graph”) i pokušava da u njemu nađe najveći indukovani podgraf. MaWISH je korišćen za poravnavanja PIN-ova kvasca, crva i vinske mušice. Graemlin je još jedan algoritam za lokalno poravnavanje mreža; on „dodeljuje ocene“ potencijalno očuvanom modulu između različitih mreža tako što izračunava logaritamski odnos (engl. “log-ratio”) verovatnoće da za taj modul postoje evolucionarna ograničenja i verovatnoće da tih ograničenja nema, pri čemu uzima u obzir filogenetske relacije između vrsta čije se mreže poravnavaju [64].

Razvijeno je nekoliko algoritama za poravnavanje mreža na globalnom nivou [185, 65, 186, 232, 130, 118, 141]. Najstariji, IsoRank, prati intuitivno zapažanje da dva čvora treba upariti (tj. poravnati, ili preslikati) samo ako se i njihovi susedi mogu upariti [185]. To se formuliše kao problem sopstvenog vektora, te se za ocenu topološke sličnosti poravnatih čvorova iz različitih mreža koriste algoritmi spektralne teorije grafova. U poravnavanje čvorova, IsoRank uključuje i ocene sličnosti sekvenci proteina u PIN mrežama dobijene pomoću algoritma BLAST [6]; konstantna težina  $\lambda$  definisana od korisnika se koristi za kontrolisanje relativnog doprinosa topoloske sličnosti, dok  $1 - \lambda$

kontrolira doprinos sličnosti sekvenci. Ovako definisana sličnost čvorova se zatim koristi u pohlepnoj strategiji (engl. “greedy strategy”) za konstruisanje poravnanja. IsoRank je korišćen za identifikovanje funkcionalnih ortologa između kvasca i vinske mušice. Kasnije je proširen za poravnavanje više od dve mreže (engl. “multiple network alignment”) [186, 130]. Dok se u ovom pristupu kombinuju informacije o topologiji i sekvencama, osmislili smo i metode poravnavanja mreža u kojima se koristi samo topologija mreža (videti niže) [118, 141]. Uz to, postoje i metode za „ispitivanje mreža“ (engl. “network querying”) u kojima se uopšte ne koristi topologija [24]. *Ispitivanje mreža* je vrsta poređenja mreža koja u PIN mreži pronalazi podmreže slične mreži koja nas interesuje, a to je obično proteinski kompleks ili biološka putanja. Zbog ograničenog prostora, ovde ne dajemo pregled metoda ispitiva-nja mreža.

Naši algoritmi za globalno poravnavanje mreža zasnovani isključivo na lokalnoj topologiji mreže, zovu se GRAAL (GRAph ALigner) [118] i H-GRAAL [141]. Pošto oni zavise samo od topologije, mogu da poravnavaju sve tipove mreža a ne samo biološke. U oba algoritma koriste se „stepeni graflleta“ definisani niže, koji obezbeđuju preciznu kvantifikaciju topološke sličnosti čvorova [140]. Ako se podsetimo da je stepen čvora broj grana koje on dodiruje i da je grana jedini grafllet sa dva čvora (grafllet  $G_0$  na slici 2), možemo definisati „stepen graflleta“ čvora  $x$  u odnosu na svaki grafllet  $G_i$  na slici 2, u smislu da  $G_i$ -stepen za  $x$  broji „koliko graflleta tipa  $G_i$  dodiruje čvor  $x$ “ [158]. Dakle, tradicionalni stepen je samo  $G_0$ -stepen. Pošto ima 30 graflleta s najviše 5 čvorova, na opisan način bi se dobio vektor sa 30 „stepena graflleta.“ Srećom, specifičnost se može povećati zahvaljujući zapažanju da svi čvorovi u graflletu nisu topološki ekvivalentni. Na primer, središnji čvor graflleta  $G_1$  topološki je različit od krajnjih čvorova graflleta  $G_1$  (slika 2). Na slici 2 prikazana su sva 73 topološki različita čvora u svakom od graflleta s najviše 5 čvorova. Svaki se zove *orbita* (tj. *orbita izomorfizma*, detalje videti u [158, 140]), a označavamo ih sa  $0, \dots, 72$ . Znači, *vektor stepena graflleta* (engl. “graphlet degree vector”, *GDV*) ili *GD-potpis* čvora  $x$  ima 73 elementa: element  $i$  predstavlja broj koji kazuje koliko puta čvor  $x$  „dodiruje“ odgovarajući grafllet u orbiti  $i$  uzimajući u obzir sve grafllete u svom susedstvu. Prema tome, on opisuje topologiju dela mreže koji okružuje čvor i „hvata“ međupovezanost čvorova do rastojanja 4 od datog čvora. Dosezanje razdaljine 4 od datog čvora možda je dovoljno da bi se skoro jednoznačno odredio položaj čvora u mreži, pošto mnoge stvarne mreže imaju svojstvo „malog sveta“ [219] (opisano u odeljku 1.3.). Pošto GRAAL pronadje čvorove sa sličnim GD-potpisima u dve mreže, on is smatra za „semena“ i siri se oko njih tražeći topoloski slične čvorove koji



okružuju semene čvorove (engl. “seed-and-extend approach”). H-GRAAL se zasniva na mađarskom algoritmu za rešavanje problema dodeljivanja (engl. “the assignment problem”), koji je algoritam kombinatorne optimizacije za pronalaženje maksimalnog težinskog sparivanja (engl. “maximum weight matching”) u bipartitnom grafu (engl. “bipartite graph”). Primenili smo ove algoritme na PIN-ove kvasca i čoveka i otkrili područja sličnosti mreža oko jedan red veličine veća od takvih područja dobijenih pomoću drugih algoritama. Uz to, koristili smo ih i za prenos znanja sa anotiranih na neanotirane delove poravnatih mreža. Štaviše, analogono korišćenju skorova dobijenih poravnavanjem genetičkih sekvenci za rekonstruisanje filogenije vrsta, koristili smo skorove poravnanja mreža da bismo takodje uspešno rekonstruisali filogeniju [118, 141]. Znači, poravnavanje mreža ima potencijal da obezbedi potpuno nov, nezavisan izvor informacija o biologiji i filogeniji.

Postoje dva razloga za razvoj metoda poravnavanja velikih bioloških mreža (kao što su PIN-ovi) koje koriste isključivo topologiju mreža. Biološkim mrežama se opisuje deo bioloških informacija, baš kao i genskim sekvencama. Pokazali smo da sekvence i topologija daju komplementaran uvid u bioloske fenomene [132]. Analogno algoritmima za poravnavanje genetičkih sekvenci, koji ne koriste biološke informacije izvan sekvenci kako bi obavili poravnavanje, upotreba bioloških informacija izvan topologije za poravnavanje mreža mogla bi nas omesti u pronalaženju informacija sadržanih samo u topologiji mreže. Tvrđimo da o biologiji najviše možemo naučiti samo nakon što razvijemo, a zatim i integrišemo pouzdane i precizne algoritme za analiziranje svakog tipa podataka zasebno [118].

## 1.5. Topologija mreže i biološke funkcije i bolesti

Povezanost topologije PIN-ova i biološke funkcije bila je predmet mnogih istraživanja. Cilj je predvideti funkciju proteina za koje trenutno ne znamo kakvu biološku funkciju obavljaju u ćeliji [177]. Slično tome, ispitivana je i uloga PIN-ova u bolestima [179]. Pomoću jedne starije metode pronađena je korelacija između povezanosti proteina, tj. stepena čvora proteina u PIN-u, i njegove neophodnosti za prezivljavanje ćelije pekarskog kvasca [96]. Međutim, ispostavilo se da je sam stepen čvora isuviše jednostavna i loša mera topologije oko čvora u mreži sa šumom i da uočena korelacija ne važi za novije PIN-ove sa manje šuma [40, 229] i da važi samo za mreže konstruisane pregledanjem literature (engl. literature-curated) [62] i manje PIN-ove dobijene ranom verzijom Y2H tehnologije [207]. Takodje, moguće je da je ovo samo posledica činjenice da su podaci o mrežama proteinskih interakcija pristrasni prema esencijalnim proteinima, tj. da proteini esencijal-

ni za život izgledaju bolje povezani u mreži samo zato što su trenutno bolje ispitani te je o njima dostupno više informacija [166]. Ispitali smo slične jednostavne korelacije između povezanosti u PIN-u i funkcije proteina [163].

To što se smatra da je visok stepen povezanosti čvora dobra mera topološkog položaja proteina u PIN-u, dovelo je do još jedne kontroverze [55, 60, 58, 59]. „Čvorišta“ (engl. “hubs”) u PIN-u proučavana su u kontekstu korelacije ekspresije gena, kolokalizacije, brzine evolucije i strukturnih perturbacija PIN-a nakon njihovog uklanjanja. Na osnovu toga, predloženo je da se pravi razlika između goreopisanih ‘raskalašnih’ čvorišta i ‘pristojnih’ čvorišta, [55, 58]. Međutim, rezultati se nisu mogli ponoviti sa skupovima PIN podataka iz literature [60, 59]. Prema tome, pored toga što je stepen slaba mera topologije mreže, kontroverza je možda i posledica odstupanja koje različite tehnike konstruisanja PIN-ova unose u topologije PIN-ova [57, 212]. Primer odstupanja koje utiče na topologiju PIN-ova jeste ono koje postoji u PIN-ovima dobijenim u AP/MS postupcima: interakcije tih PIN-ova se obično modeluju ili pomoću „matrice“ (pretpostavlja se da svi proteini povezani u istaloženom proteinskom kompleksu stupaju u direktne interakcije), ili pomoću modela „prečage, žbice“ (engl. “spoke”; pretpostavlja se da protein koji služi kao „mamac” za istaloženi kompleks stupa u interakcije sa svim proteinima „žrtvama”, ali da nema drugih interakcija u kompleksu) [10, 223], a pokazano je da izbor modela utiče na topologiju PIN-ova [78].

Zbog pomenutih kontroverzi vezanih za korišćenje previše pojednostavljenih mera za topologije mreže, kao što su stepeni čvora, razvijen je niz sofisticiranijih metoda teorije grafova. Neke se zasnivaju na pretpostavci da je verovatnije da će proteini koji su bliže jedni drugima u PIN-u imati slične funkcije [87, 174, 30], u drugima se pokušava da se minimizuje broj protein-protein (PP) interakcija između različitih funkcionalnih kategorija u PIN-u [211, 102], ili se koriste metode bazirane na preseku i protoku kroz mreže (“network-flow” metode bazirane na minimalnom preseku i maksimalnom protoku) [146]. Uz to, ispitivana je i funkcionalna homogenost grupa proteina koje ispoljavaju neku vrstu „klasterovanja” ili „koherencije“ u PIN-u [11, 163, 175, 111, 117, 61, 100]. Vektori stepena graflata, opisani u odeljku 1.4., takođe su korišćeni za izolovanje topoloških položaja proteina u PIN-ovima i njihovo povezivanje sa funkcijama proteina i učešću u bolesti [140, 53, 139]. Detalje videti u narednim poglavljima.

Slično tome, ispitivani su ljudski PIN-ovi u potrazi za vezama između topologije i bolesti. I u ovom slučaju, kada su samo stepeni čvorova korišćeni za merenje topologije, zapažena su sledeća neslaganja: neke grupe su saopštile da geni (i proteini) koji učestvuju u bolestima obično imaju visoke stepene u PIN-ovima [214, 98], dok su drugi osporavali taj zaključak [56]. Međutim,

opšte zapažanje je da su proteini koji izazivaju bolesti bliže jedni drugima i da su centralno pozicionirani u PIN-u [214, 98, 56]. Ali, ti rezultati možda nisu korektni pošto je moguće da proteini koji izazivaju bolesti ispoljavaju ta svojstva samo zato što su bolje proučavani od proteina koji ne uzrokuju bolesti [179]. Strožija merenja topologije mreža, kao što su ona pomoću vektora stepena graflata, takođe su sprovedena, i otkriveno je da se topološke sličnosti poznatih regulatora melanogeneze u ljudskom PIN-u mogu koristiti za predviđanje novih regulatora melanogeneze koji učestvuju u pojavi kancera [139]. Uz to, simuliranjem prostiranja apstraktnog funkcionalnog „toka“ (gde „funkcija“ znači poznato učešće nekog proteina u bolesti) od proteina koji izazivaju bolesti do njihovih suseda u PIN-u, korišćeno je za ocenjivanje jačine povezanosti proteina s bolešću i proteinskim kompleksima [208]; ova metoda nema nikakve veze s dinamikom mreže – to je metoda primenjena na (statički) PIN da bi se ocenilo i predvidelo učešće proteina u nekoj bolesti. Predložen je i pristup u kome se kombinuju različite vrste bioloških podataka s PIN-ovima kako bi se predvideli geni kancera [7].

Opšti zaključak je sledeći: veza između topologije mreže i biološke funkcije s jedne strane i učešća u bolesti s druge, daleko je od toga da je slučajna, mada u ovm trenutku nismo u stanju da matematički potpuno opišemo te veze.

## 1.6. Prognoza

Novonastala istraživačka oblast bioloških mreža već je u previranju. Ona okuplja istraživače iz različitih disciplina koji su na različitim nivoima razumevanja biologije i kvantitativnih merenja. Zbog toga se daje prednost upotrebi jednostavnih mera za topologiju mreža, kao što su stepeni čvorova, i modelovanju jednostavnih bioloških fenomena. Međutim, kao što je gore pokazano, takvi postupci daju neuverljive ili kontroverzne rezultate, što jasno ukazuje na potrebu razvijanja boljih algoritamskih i matematičkih alata. Obeshrabreni takvim kontroverzama, neke podzajednice su počele da razvijaju otpor prema „preterano složenim“ metodama i novim modelima. Takvi stavovi bi možda mogli da vode ka pojavi previše pojednostavljenih doktrina koje bi onda mogle da nanesu dalju štetu ovoj narastajućoj istraživačkoj oblasti. Na primer, naširoko objavljivana primena jednostavnih tehnika izračunavanja, kao što su raspodele stepena, na rane PIN-ove s mnogo šuma, doprinelo je promovisanju modelovanja kompleksnih bioloških mreža zasnovanog na mrežama bez skale [108], za koje se kasnije pokazalo da predstavljaju previše pojednostavljen model za takve mreže.

Drugi primer je genom-centrični pogled na biološke sisteme. Analize

podataka iz genskih sekvenci svakako predstavljaju revoluciju u našem viđenju biologije. Međutim, one nam nisu omogućile da potpuno razumemo biološke sisteme, pa treba da budemo otvoreni za nove naučne horizonte. Posebno treba istaći sledeće: mada su PIN-ovi novi oblik podataka koji su trenutno puni šuma i nekompletni, i mada je većina matematičkih metoda koja se trenutno primenjuje na njih prilično primitivna, već postoje dokazi koji potvrđuju da oni mogu otkriti biološke informacije koje se ne bi mogle izvući iz genskih sekvenci – bar ne uz pomoć trenutno raspoloživih alata za analizu sekvenci. Uprkos takvim rezultatima, neki članovi zajednice i dalje dovode u pitanje vrednost podataka iz mreža iznoseći nedokazane tvrdnje koje osporavaju povezanost topologije mreža i biološke funkcije. Na nedavnom sastanku, to je dovelo do javno postavljenog pitanja da li bi zajednica „trebalo da nastavi analiziranje podataka iz PIN mreža.“ Izlišno je reći da postoji opasnost da takva pitanja dovedu do naučnog cenzurisanja koje bi ta naučna oblast već mogla početi da oseća. Na primer, istraživanja koja pokazuju da podaci iz PIN-ova otkrivaju informacije koje se ne mogu dobiti iz genskih sekvenci mogu se smatrati pogrešnim pošto se ne slažu sa uvreženim verovanjima o svemoći genetičkih sekvenci. S druge strane, ukoliko su istraživanja zasnovana na PIN-ovima *u skladu* sa onim što nam govore sekvence, onda se ona često smatraju beskorisnim, jer samo potvrđuju ono što se može dobiti analizom sekvenci. Prema tome, PIN analize će verovatno gubiti u takvoj nepoštenoj igri. Štaviše, takvi stavovi mogu negativno uticati na raspoloživost finansijskih resursa neophodnih za kontinualno prikupljanje podataka, kao i za prateći razvoj pouzdanih metoda izračunavanja i modelovanja. I analiza genoma i analiza topologije mreža glavni su domeni savremene bioinformatike i njihovo nezdravo sukobljavanje, do kojeg ponekad dolazi na sastancima bioinformatičara, slično je bavljenju nevažnim temama u lokalnoj politici u trenutku kada postoji pretnja spolja. U suštini, rasprava o analizi genoma ili PIN-ova kada se gleda iz šire perspektive novih izazova u biologiji sistema sasvim je marginalna.

Na kraju, postoji i izvestan sukob unutar zajednice koja se bavi izračunavanjima. Mada su se statističke analize podataka pokazale vrednim, takvi pristupi ne daju mehanicističke opise načina na koje stvari funkcionišu. Prema tome, bez obzira na prevalenciju i lakoću upotrebe statističkih alata, trebalo bi da nastavimo da tražimo opisne modele pomoću kojih možemo razumeti i reprodukovati dati fenomen.

Ovo poglavlje je bazirano na članku:

N. Pržulj, “Protein-protein interactions: Making sense of networks via graph-theoretic modeling,” *Bioessays*, DOI 10.1002/bies.20100044, February 2011.