



UNIVERSITATEA BABEȘ-BOLYAI
BABEȘ-BOLYAI TUDOMÁNYEGYETEM
BABEȘ-BOLYAI UNIVERSITÁT

TRADITIO ET EXCELLENTIA

Metode de fizică statistică pentru studiul rețelelor complexe

disertație de doctorat prezentată

de

István Papp

la

Departamentul de Fizică

Conducător științific

Professor Dr. Zoltán Néda

Babeș-Bolyai University

Cluj-Napoca, Romania

2020

Metode de fizică statistică pentru studiul rețelelor complexe

Abstract

Calculatoarele au trecut printr-o evoluție rapidă în ultimele decenii, ceea ce a adus la progrese în științele bazate pe date. Pe măsură ce prelucrarea datelor a devenit mai ușoară, bazele de date s-au dezvoltat vertiginos și au devenit accesibile pe Internet. Datele de baza referitoare la rețele au crescut odată cu cantitatea informațiilor colectate și au schimbat complet modul în care comunitatea științifică consideră rețele. În prezent multe sisteme complexe sunt descriși prin rețele, în timp ce în aproape toate domeniile științei de la biologie la informatică, inginerie, economie, politică etc., apar întrebări legate de găsirea structurii comunităților pe aceste grafuri. Definiția tradițională a acestor comunități se bazează pe conectivitatea nodurilor, membrii unei comunități au mai multe conexiuni interne, între ele decât cu restul rețelei. În cadrul lucrării, în aceasta tematică s-a implementat un algoritm pentru o astfel de partiționare bazat pe conceptul de diagrame Voronoi, utilizate în principal pentru împărțirea spațiului geometric. În cadrul acestei metode s-a definit o metrică folosind coeficientul de Edge Clustering și s-a introdus o măsură de densitate locală bazată pe aceste distanțe pentru a identifica nodurile generatoare Voronoi. S-a dezvoltat metoda inițială prin selectarea generatoarelor la întâmplare și efectuarea partiționării Voronoi de mai multe ori cu diferite seturi de generator. În partea a doua a tezei, s-a studiat mobilitatea umană explorând universalități între diferite moduri de mobilități și s-a studiat dinamica transferului de date pe Internet. Rezultatele au sugerat o tendință în care viteza de propagare crește monoton în forma unei legi de putere cu distanță. S-a arătat că originea acestor tendințe nu se datorează numai structurii rețelelor, existând și întâzieri pe noduri individuale, inclusiv pe punctele de pornire și de sosire. Pentru o mai bună înțelegere topologiei de conectare al Internetului am introdus un model care era capabil să reproducă atât structura, cât și caracteristicile dinamice de scalare observate. Teza conține trei capitole mari: o introducere în domeniu; o descriere teoretică și matematică a partiționării de rețele; iar ultimul capitol conține experimente, modele și studiul universalităților în rețelele de mobilitate și Internet.

Cuprinsul tezei de doctorat

1	De ce este interesant acest domeniu pentru un fizician?	1
2	Detectarea comunității în rețelele complexe	9
2.1	Detectarea cu diagrame graf-Voronoi	9
2.1.1	Ilustrarea metodei în spațiul geometric	10
2.1.2	Generalizarea metodei pe rețele	12
2.1.3	Rezultate	17
2.2	Detectarea cu diagrame graf-Voronoi stohastice	21
2.2.1	Diagrame Voronoi stohastice	22
2.2.2	Abordare analitică pentru rețele non-ierarhice mari	25
2.2.3	Case 1: Modularitate extremă	26
2.2.4	Case 2: Densitate de conexiune intermodulară non-infinitesimală	32
2.2.5	Complexitate vs. precizie	35
2.2.6	Metoda de creștere a contrastului	37
2.2.7	Rețele reale	39
2.3	Prezentare generală scurtă	45
3	Scalare în rețele reale	49
3.1	Scalare în dinamica mobilităților	49
3.2	Scalare pe Internet, experiment și model	56
3.2.1	Experimentele	56
3.2.2	Modelul	58
3.2.3	Modelul vs. realitate. Rezultate și discuții	63
4	Concluzii	69
4.1	Detectarea de comunități	69
4.2	Legi de Scalare	71
	Addendum A Rețele de referință	73

Addendum B	Găsirea generatoarelor	75
Addendum C	Comparația cu alte metode de detectare	77
Addendum D	Rețele din lumea reală	79
Addendum E	Datele și metode de binning	83
5	Publicații	87
5.1	Publicații relevante pentru teză	87
5.1.1	Lucrări științifice	87
5.1.2	Prezentări la conferințe	88
5.2	Alte publicații	88
5.2.1	Lucrări științifice	88
5.2.2	Prezentări la conferințe	88
5.2.3	Brevete	89
	Bibliografie	101

Cuprins

1	Introducere	1
2	Detectarea comunităților în rețele complexe	3
2.1	Detectare cu diagrame Voronoi	3
2.2	Detectare cu diagrame Voronoi stohastice	6
3	Scalări în rețele reale	11
3.1	Scalare în dinamica mobilităților	11
3.2	Scalări în rețeaua Internetului, experiment și model	14
3.2.1	Modelul de Internet și rezultatele	14
4	Concluzii	18
4.1	Detectarea comunităților în rețele complexe	18
4.2	Legi de scalare în rețele reale	19
5	Lista de publicații	20
	Bibliografie selectată	23

Cuvinte cheie

rețele complexe, rețele aleatoare, rețele scale-free, topologie de rețea, sisteme complexe, detectarea comunităților în rețele, diagrame Voronoi, legi de scalare, universalități, Internet

1

Introducere

Științele sociale, teoria informației, tehnologia, biologia, neuro-știința etc., sunt domenii care studiază sisteme ce pot fi reprezentate cu rețele iar analiza rețelilor a devenit un instrument util în înțelegerea acestora [1, 2].

În aproape toate sistemele reprezentate cu grafuri apare cel puțin o problemă, unde găsierea structurii comunității rețelei joacă un rol cheie la înțelegerea acesteia. Un exemplu imediat este clusterizarea publicațiilor într-o rețea de citări cu domenii suprapuse.

De obicei comunitățile sunt definite ca noduri conectate mai compact în grupul lor decât cu celelalte noduri din rețea. Dificultatea la identificarea oricărei structuri de comunități ar putea fi în definirea acestora, întrucât acesta este una calitativă, și încă nu există o definiție matematică acceptată [3]. Deși există o mare varietate de definiții și metode de detectare a comunităților [3], combinarea definițiilor matematice semnificative cu algoritmi eficienți rămâne o problemă. Probleme similare apar și în extragerea datelor, recunoașterea formelor, învățarea automată și analiza statistică a datelor [4, 5]. Aici însă problemele sunt definite în spații metrice continue, ceea ce simplifică problema. Diagramele Voronoi [6] sunt utilizate în mod obișnuit pentru a împărți spațiul metric în subsecțiuni, numite celule Voronoi.

Rețelele încorporate în spațiu metric sunt foarte importante, nu numai pentru clusterizare eficientă, dar ne înconjoară în natură. Chiar și mobilitatea umană sau canalele noastre de comunicare, de exemplu Internetul, este o rețea încorporată în spațiul metric. Din experiența noastră zilnică știm, că timpul de călătorie nu crește liniar cu distanța de călătorie [7–9]. Cauza acestui lucru este complexă și include diferite efecte. Pentru a acoperi distanțe mari folosim autostrăzi, dar timpul călătoriei de multe ori nu este influențat de distanță ci de timpul pierdut cu intrarea și ieșirea din orașe. Prin urmare, timpul de călătorie între orașe apropiate poate fi și ea mare. O situație similară este observată în traficul aerian. O mare parte din timp trece cu decolare, aterizare și parcare. In-

ternetul la nivel de router fiind o rețea complexă asemănătoare încorporată în spațiul geografic, pe lângă proprietățile sale de scalare topologică (distribuție fără scalare) [10, 11], prezintă o scalare dinamică, similară cu mobilitatea umană. În lucrare se studiază aceste fenomene mai profund, și căutăm cauzele vizibile ale acestora folosind o varietate de baze de date online și urmărirea GPS. Pentru studiul dinamicii pachetelor de date pe Internet folosim metodele bazate pe protocolul mesajelor de control: Ping și Traceroute.

Pentru a explica legile de scalare observate și alte proprietăți topologice măsurabile ale Internetului, propunem un model simplu. Un astfel de model trebuie să se bazeze pe ipoteze realiste cu privire la procesul de cablare și trebuie să reproducă proprietățile topologice măsurate, inclusiv scalarea vitezei de comunicare în funcție de distanță.

Teza conține două capitole principale. În primul capitol este prezentată descrierea matematică și aplicarea metodei noastre de identificare a comunităților în rețele, argumentând avantajele rețelilor încorporate în spațiu. Al doilea capitol conține experimente și un model pentru proprietățile de scalare a rețelelor de mobilitate și a Internetului.

2

Detectarea comunităților în rețele complexe

2.1 Detectare cu diagrame Voronoi

Pentru a prezenta tehnica noastră de clusterizare, demonstrăm mai întâi aspectele sale esențiale cu o problemă similară definită în spațiul Euclidian 2D (figura 2.1(a)).

Să considerăm $G = (V, E)$ un graf direcționat ponderat cu V din N noduri, și un set E din M legături. $l(\mu, \nu) > 0$ semnifică greutatea unei legături care conectează nodul μ și ν . Lungimea unui traseu în graf este obținută prin adăugarea greutăților legăturilor. Notăm distanța care leagă două noduri ν_i și ν_j cu $d(\nu_i, \nu_j)$, acesta înseamnă calea cea mai scurtă dintre ele. Definiția această a lungimei asigură încadrarea rețelei într-un spațiu metric. Selecția mai simplă de $l(\mu, \nu)$ este 0 când μ și ν nu sunt conectate, iar 1 când sunt conectate direct. Considerăm un grup de $S \equiv (\gamma_1, \gamma_2, \dots, \gamma_g) \subset V$ noduri generatoare. Partiționarea Voronoi al rețelei G respectând S va fi împărțirea a lui V în grupuri de noduri $V_1, V_2, \dots, V_g \subset V$. Fiecare grup (celula Voronoi) aparține unui generator și satisface următoarele: i) Rețeaua G conține toate celulele Voronoi fără suprapunere; ii) Toate nodurile dintr-o celulă sunt mai aproape de generatorul lor decât de oricare generator din rețea. Caracteristicile matematice ale diagramelor Voronoi în rețele sunt descrise în [6] împreună cu complexitatea lor computațională.

Pentru a folosi aproximația aceasta geometrică, avem nevoie de o definiție adecvată pentru măsurarea distanței, care transformă apartenența nodurilor în segregarea spațiului metric. Am ales una dintre cele mai simple măsuri: coeficientul de clusterizare a legăturilor (ECC) propusă în

publicația [12]. Legătura care conectează vertexul i și vertexul j are ECC definit ca:

$$C_{i,j} = \frac{z_{i,j}}{\min[(k_i - 1), (k_j - 1)]} \quad (2.1)$$

unde k_i, k_j sunt numărul de conexiuni acestor două noduri. $z_{i,j}$ este numărul triunghiurilor din care face parte legătura și $\min[(k_i - 1), (k_j - 1)]$ este numărul triunghiurilor posibile la care ar putea aparține. Cu cât ECC este mai scăzut, cu atât este mai probabil ca nodurile legate aparțin în comunități diferite. Prin urmare, în metoda noastră de partiționare, specificăm lungimea (greutatea) conexiunii ca valoarea inversă a lui ECC. Lungimea oricărei traseu din rețea este suma greutateților de-a lungul căii, iar lungimea celei mai scurte căi definim ca distanța dintre noduri.

Obiectivul principal este alegerea unui nod privilegiat în fiecare comunitate. Selectăm nodurile generatoare folosind densitatea lor locală relativă [3], definit ca:

$$\rho_i = \frac{m}{m + k} \quad (2.2)$$

unde m este gradul intern (numărul conexiunilor de intrare) în subrețea care conține primii vecini al nodului i , iar k este gradul extern al vecinătății (numărul legăturilor de ieșire). Această densitate este mai mare pentru nodurile aflate „în centrul” comunităților, așa cum este ilustrat în figura 2.1(b). La fel ca în spațiul 2D, generatoare vor fi nodurile cu densități maxime într-o rază r . În general, complexitatea de detectare a nodurilor generatoare rămâne mult sub $O(Mg)$ (pentru g generatoare Voronoi), iar partiționarea are o complexitate de $O(N \log N)$ [6].

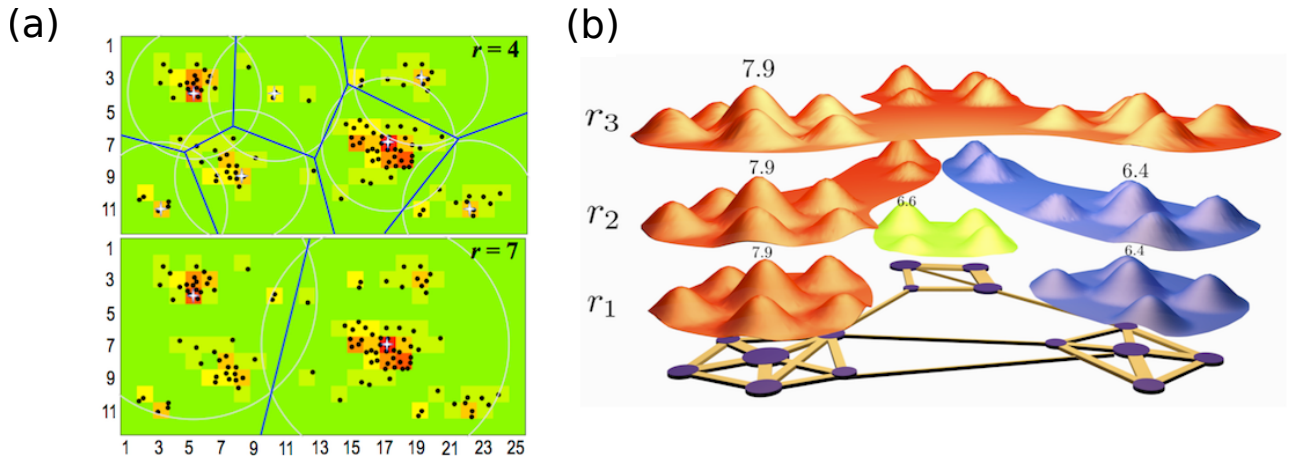


Figura 2.1: Localizarea centrelor Voronoi. (a) Pentru a găsi cea mai bună clusterizare a punctelor negre, împărțim planul în 12×25 pătrate și estimăm densitatea locală (numărul de puncte) în fiecare pătrat. Pătratele sunt colorate de la verde (0 puncte) la roșu (mari de 6 puncte) în funcție de densitatea lor. Pătratul cu cea mai mare densitate într-o rază de r (cercuri gri, $r = 4$ în sus și $r = 7$ în unități pătrate), devine un pătrat generator (marcată cu stea albă). Partiționarea Voronoi este ilustrat cu linii albastre. (b) Ilustrarea implementării algoritmului în rețele: fiecare nod va avea o densitate locală, indicată de mărimea nodurilor în “muntele” comunității. Lungimile conexiunilor sunt invers proporționale cu coeficientul de clusterizare al conexiunilor (ilustrat cu grosimea liniilor). Mărimea comunităților crește cu mărimea razei r ($r_1 < r_2 < r_3$).

Așa cum era subliniat în articolul de referință [3], este necesar ca metodele de detectare să fie testate atât pe rețele de referință (generate cu comunități predeterminate) cât și pe rețele reale.

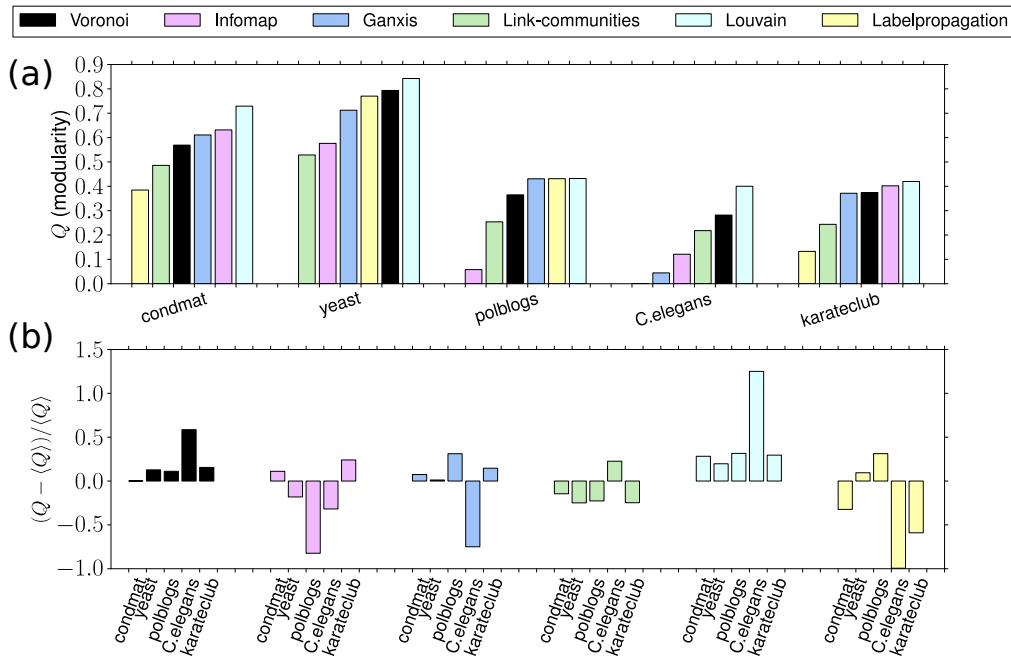


Figura 2.2: Testarea pe rețele reale. Modularitate obținută pe toate cele cinci rețele reale cu metode diferite (specificate în legendă). (a) Modularități Q în ordine crescătoare pe fiecare rețea testată. (b) Eroarea relativă $(Q - \langle Q \rangle) / \langle Q \rangle$, aici $\langle Q \rangle$ reprezintă media modularităților rezultat prin fiecare algoritm aplicat pe rețele considerate.

Detectarea structurii de comunități în grafuri din lumea reală nu este însă trivial. În plus, eficiența oricărei metode este mai dificil de evaluat, deoarece informațiile despre adevărul de bază nu sunt disponibile (adică comunitățile nu sunt cunoscute) ca în cazul rețelelor de referință. Prin urmare, noile algoritmi sunt testați comparativ cu metodele cunoscute și acceptate printr-o măsură de calitate. În cazul acesta am ales modularitatea. Modularitatea este definită după cum urmează: $Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(V_i, V_j)$, aici sumarea este pe toate perechile de noduri. A este matricea adiacență, m numărul total de legături în rețea și P_{ij} reprezintă numărul legăturilor așteptate între nodurile i și j în modelul null. Modelul nul în cazul nostru este, modelul grafului aleatoriu. Funcția δ returnează 1 dacă nodurile i și j se află în aceeași comunitate adică $V_i = V_j$, în caz contrar returnează 0 [13]. Am evaluat algoritmul pe mai multe rețele reale, care sunt utilizate în mod regulat în literatură. Structural și după origine acestea sunt foarte diferite:

1. Rețeaua de prieteni din clubul karate studiat de W.Zachary, conține 34 de membri conectați cu 78 de legături [14].
2. Rețeaua de neuroni din nematodul *Caenorhabditis elegans* [15, 16] conținând 297 noduri și 2359 conexiuni.
3. Rețeaua de interacțiuni proteină-proteină a drojdiei [17] cuprinzând 1845 de noduri conectate prin 4405 de legături.

4. O versiune revizuită a rețelei de colaborare între oamenii de știință din arhiva www.arxiv.org. Această rețea este formată din 39576 noduri și 175692 conexiuni, construit după articole publicate în arhivă din perioada 1 ianuarie 1995 - 31 martie 2005 [18].
5. Conexiuni între blogurile politice americane [19] care include 1223 noduri și 19087 de legături.

Am comparat algoritmul nostru cu cinci metode frecvent utilizate în literatură (pentru mai multe informații, a se vedea Addendumul tezei): 1) algoritmul Louvain optimizează modularitatea [20]; 2) the label propagation algorithm (LPA) [21]; 3) GANXiS sau SLPA (speaker-listener label propagation algorithm) [22, 23]; 4) link-communitiy detection [24] și 5) Infomap (IM) [25, 26]. Metoda noastră ocupă locul doi din punct de vedere al modularității, modularitatea rezultată fiind întotdeauna peste medie. Singurul algoritmul care obține performanțe mai bune este Louvain, era de așteptat, deoarece optimizează același funcție de calitate pe care folosim la evaluarea performanței.

2.2 Detectare cu diagrame Voronoi stohastice

Pentru o anumită rețea nedirecționată, cu noduri N și conexiuni M , neglijăm selecția generatoarelor Voronoi. În schimb, folosim un set de celule Voronoi, unde fiecare set este obținut de la generatoare alese aleator [27] respectând următoarele: i) Selectăm la întâmplare un număr de noduri

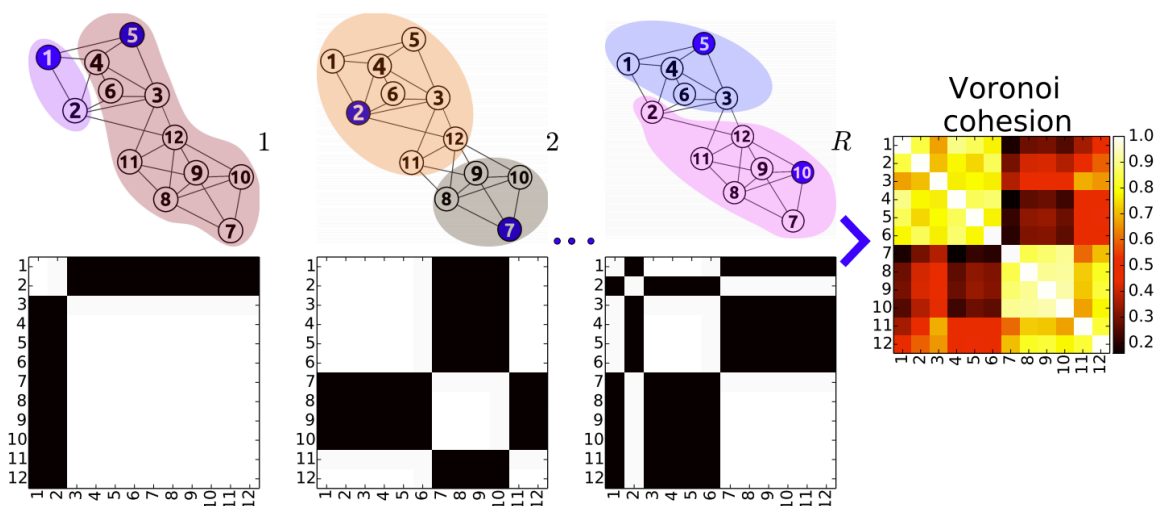


Figura 2.3: Idea principală pentru diagrame Voronoi stohastice. Figurile arată modul în care harta de coeziune Voronoi este calculată după numărul de R partiționări. Fiecare sub-figură arată o realizare a matricei de colocare binară generată cu 2 noduri generatoare alese aleatoriu de fiecare dată: de exemplu (1, 5), (2, 7), ..., (5, 10). Valorile matricei de colocare sunt 1 = *alb* pentru nodurile din aceeași celulă și 0 = *negru* dacă nodurile sunt în celule diferite. Media peste toate R matrice rezultă harta de coeziune, cu alte cuvinte, probabilitățile ca nodurile să fie în aceeași comunitate.

g din rețea și le folosim ca generatoare Voronoi pentru a efectua o partiționare; ii) Obținem matricea de coeziune Voronoi (probabilitățile de co-locare) după medierea peste R matrici de co-locare (figura 2.3).

Pentru testarea teoriei, utilizăm rețele aleatoare, nedirecționate cu $N \rightarrow \infty$ noduri organizate în m comunități de tip Erdős-Rényi (ER). Folosim comunități conectate fără suprapunere, de dimensiuni $N_i \equiv \alpha_i N$, $0 < \alpha_i < 1$ și matricea de densitate a conexiunii $q_{ij} = M_{ij}/(N_i N_j)$, $i, j = \overline{1, m}$ caracterizează conectivitatea lor. Acesta înseamnă probabilitatea conexiunii între module i și j . M_{ij} reprezintă numărul de conexiuni care conectează cele două module, cu o dependență $\mathcal{O}(N^2)$ de dimensiunea rețelei. Notăm cu q_i elementele diagonale ale matricei și ne vom referi la ele ca densități de conexiune intra-modulare, în timp ce elementele off-diagonale vor fi denumite ca densități de conexiune inter-modulare (poduri). Numărul nodurilor din modulul i care formează un „pod” cu modulul j este B_{ij} . Să considerăm evenimentele: i) X_{ij} - două noduri din comunitățile i și j , $i, j = \overline{1, m}$, sunt alocate în aceeași celulă Voronoi; ii) $G_{n_1 n_2 \dots n_m}$, $\sum_{i=1}^m n_i = g$ - nodurile generatoare sunt împărțite între m comunități, astfel încât n_i generatoare vor fi în modulul i . $\{G_{n_1 n_2 \dots n_m}\}$ este un set complet din C_{g+m-1}^{m-1} evenimente. Acesta este echivalent cu numărul realizărilor pentru separarea unui lanț liniar de g bile egale cu $m - 1$ bariere poziționate aleator. C_n^k indică coeficientul binomial pentru combinațiile k ale n evenimente. Corespunzător:

$$X_{ij} = \sum_{n_1 n_2 \dots n_g} X_{ij} \cdot G_{n_1 n_2 \dots n_g} \cdot \quad (2.3)$$

Atunci probabilitatea unui vertex din comunitatea i și unul din comunitatea j aparținând la aceeași celulă Voronoi (sau coeziunea Voronoi) este:

$$c_{ij} \equiv P(X_{ij}) = \sum_{n_1 n_2 \dots n_g} P(X_{ij} | G_{n_1 n_2 \dots n_g}) P(G_{n_1 n_2 \dots n_g}),$$

care se poate scrie în forma:

$$\mathbf{c} = \mathbf{V} \cdot \mathbf{g}, \quad (2.4)$$

unde \mathbf{c} , \mathbf{V} și \mathbf{g} după cum urmează sunt matrici de mărime $N(N+1)/2 \times \mathbf{1}$, $N(N+1)/2 \times C_{g+m-1}^{m-1}$ și $C_{g+m-1}^{m-1} \times \mathbf{1}$.

Să considerăm două module echivalente, ambele cu dimensiunea $N \gg 1$, densitatea conexiunii q , și densitatea podului b . În continuare prezentăm câteva cantități semnificative ale unei perechi de noduri în funcția de locul lor (în modulul 1. sau 2.). Primul tabel sintetizează distribuția de probabilități ale distanțelor comune. Presupunând că ambele noduri sunt generatoare, al doilea tabel conține dimensiunile relative ale celulelor Voronoi potrivite în toate module:

loc. în mod.	$d=1$		modulul 1		modulul 2		loc. în mod.
	$d=1$	$d=2$	gen.1	gen.2	gen.1	gen.2	
1,1	q	$1-q$	1/2	1/2	1/2	1/2	1,1
1,2	b	$1-b$	s	$1-s$	$1-s$	s	1,2
2,2	q	$1-q$	1/2	1/2	1/2	1/2	2,2

unde s este dimensiunea relativă a celulei Voronoi 1 din modulul 1, când vertexul generatorului 2 este situat în modulul 2. Valoarea lui s poate fi calculată prin însumarea următoarelor contribuții (figura 2.4(a)):

1. fracția bq din modulul 1 are conexiuni directe la ambele generatoare și este împărțită uniform între cele două celule Voronoi;
2. fracția $q(1-b)$ are conexiuni directe doar la generatorul 1, prin urmă contribuie total la celula 1;
3. fracția $(1-q)b$ are conexiune directă la generatorul 2, prin urmare nu contribuie la celula 1;
4. fracția $(1-b)(1-q)$ nu este conectat direct la ambele generatoare. Acesta este împărțit între cele două generatoare proporțional cu numărul de căi mai scurte. Nodul „modus operandi” pentru generatorul 1 poate fi din modulul 1 cu o probabilitate de q^2 și cu b^2 de la modulul 2. Valorile similare pentru generatorul 2 sunt ambii bq , prin urmare, o fracțiune de $(q^2 + b^2)/(q + b)^2$ din acest domeniu corespunde la celula 1.

Adăugarea contribuțiilor rezultă:

$$s(q, b) = \frac{qb}{2} + \frac{1-b}{(q+b)^2} [q^2 + b^2 + 2q^2b] . \quad (2.5)$$

Construind matricea \mathbf{c} din ecuația (2.4):

$$\mathbf{v} = \begin{pmatrix} 1/2 & 1 - 2s(1-s) & 1/2 \\ 1/2 & 2s(1-s) & 1/2 \\ 1/2 & 1 - 2s(1-s) & 1/2 \end{pmatrix}, \quad \mathbf{g} = \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix} .$$

Prin urmare

$$c_{11} = c_{22} = \frac{3}{4} - s(1-s), \quad (2.6)$$

$$c_{12} = \frac{1}{4} + s(1-s). \quad (2.7)$$

unde contrastul dependent de densitatea de conexiune este

$$\gamma = \frac{1}{2} [1 - 4s(1-s)]. \quad (2.8)$$

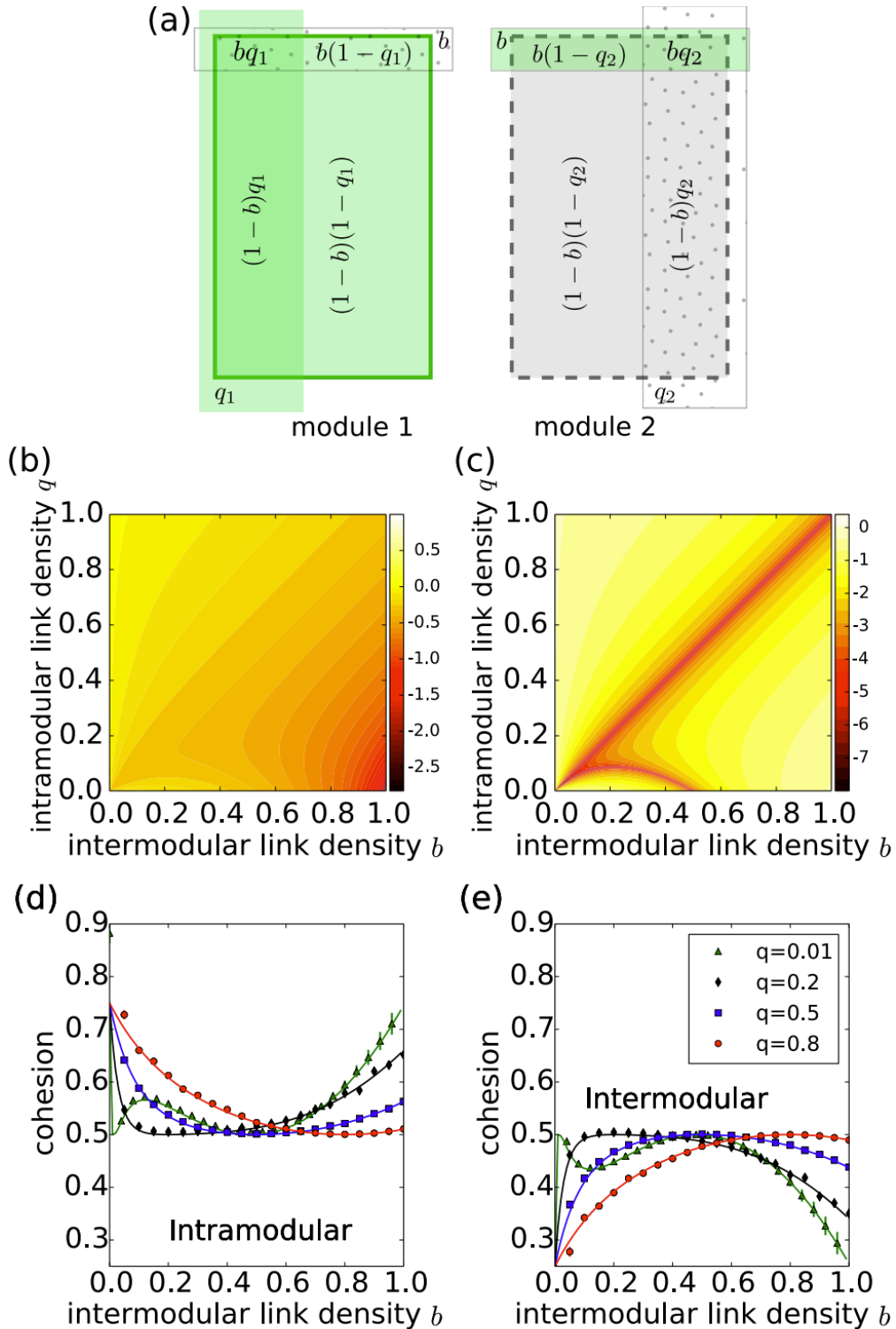


Figura 2.4: Densități de conexiune inter-modulare. (a) Ilustrarea cum două regiuni Voronoi împărțesc două moduli de tip ER cu dimensiuni egale având densitate de conexiune finită. (b) Arată modul în care densitățile de conexiune intra-modulare, $q_1 = q_2 = q$, și inter-modular, b , influențez dimensiunea celulei 1 în modulul 1 (descriș cu ecuația (2.5)). (c) Arată contrastul corespunzător din ecuația (2.8). (d, e) Arată coeziunile Voronoi intra- și inter-modulare în funcția de densitate a conexiunii. Rezultatele analitice din ecuațiile (2.6) și (2.7) pentru mai multe densități de conexiune intra-modulare sunt prezentate în comparație cu rezultate de simulări obținute din 5000 de partiționări într-un graf cu $N = 2 \times 800$ de noduri. Aceasta înseamnă 10 structuri diferite cu 500 de seturi de noduri generatoare.

Dacă $b = q$ densitățile de conexiuni intra și intermodulare se egalează, $\gamma = 0$, înseamnă că rețeaua nu mai este modulară. Contrastul maxim este $\gamma = 1/2$, când dimensiunea podului este neglijabil. Caracteristica interesantă a metodei este că rezultatele analitice poate fi confirmate în rețele cu dimensiuni reduse (figura 2.4).

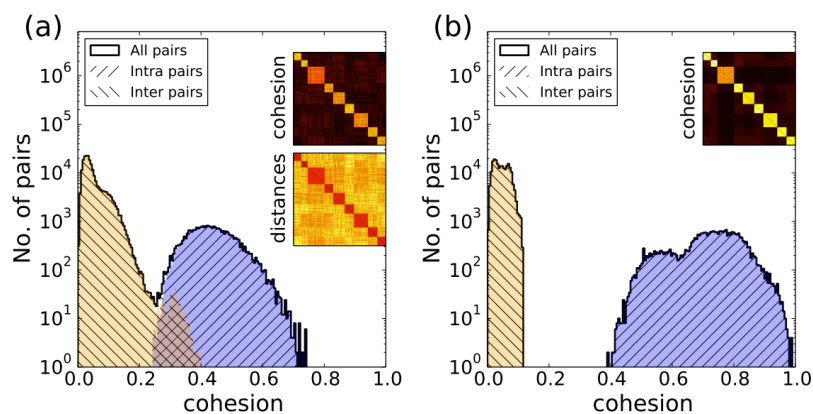


Figura 2.5: Matricea de coeziune în rețele de referință. (a) Histograma de coeziune pentru fiecare pereche de noduri din harta de coeziune a unei rețele de referință format din $N = 500$ noduri, $M = 5000$ conexiuni, și $m = 9$ module [28, 29]. Pentru a calcula coeziunea, $g = 15$ generatoare și $R = 3000$ repetiții erau implicate. Perechile de noduri intra- și inter-modulare sunt colorate după adevărul de bază (comunitățile cunoscute). Subfigurile arată matricile de coeziune și de distanță dintre noduri, ordonate după adevărul de bază, și aici valorile mai mari sunt mai aproape de linia diagonală principală. (b) Histograma de coeziune calculată pentru același rețea, dar cu metoda de creștere a contrastului (mai multe detalii în secțiunea 2.2.6 a tezei).

Structura comunităților în rețele poate fi dedusă din matricea de coeziune. Cea mai simplă metodă este de a pune un prag în histograma de coeziune (figura 2.5(a)). În figura 2.5(a) arătăm distribuția valorilor din matricea de coeziune Voronoi pentru toate perechile de noduri. Pentru a optima partiționarea utilizăm informațiile matricei de coeziune. După fiecare 200 de repetiții, mutăm conexiunile nodurilor necorelate între nodurile neconectate dar corelate (figura 2.5(b)). Această metodă crește diferența între nodurile intra și intermodulare păstrând structura comunităților.

Comunitățile pot fi detectate astfel: 1. inițial fiecare nod primește o etichetă de comunitate separată; 2. eticheta unui nod va fi schimbată în eticheta altui nod, care are o etichetă neschimbată și are o valoare de coeziune similară sub pragul definit. Sporirea contrastului rar funcționează în special la rețelele reale, dar poate îmbunătăți vizibil identificarea comunităților. Analiza detaliată cu rețele reale poate fi găsit în teza.

3

Scalări în rețele reale

3.1 Scalare în dinamica mobilităților

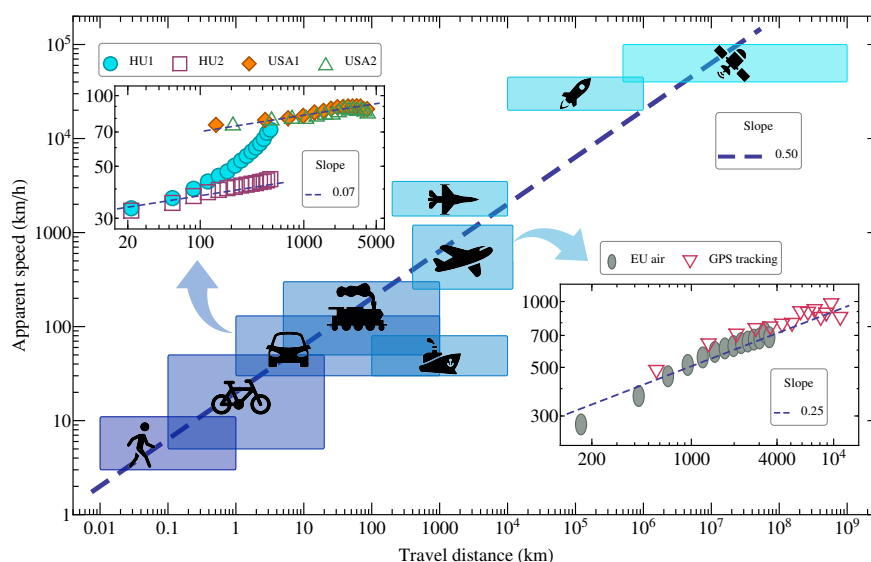


Figura 3.1: Viteza de călătorie aparentă pentru toate modurile de transport (estimat din timpul de călătorie pe linia geodetică) în funcție de distanță de deplasare. Dreptunghiurile indică intervalele de viteză și distanță la care este utilizat modul specificat. Linia punctată reprezintă o tendință de scalare cu exponent 0,5. Cele două figuri interioare prezintă câteva rezultate mediate pentru cele mai populare moduri de transport: călătoriile cu mașina și cu avionul. Liniile punctate nu sunt regresii, ci indică tendințele de scalare cu exponenții specificați numai pentru ghidarea ochilor. Axele sunt logaritmice.

Pentru toate modurile de deplasare umană, de la pietoni până la transportul cosmic, studiem

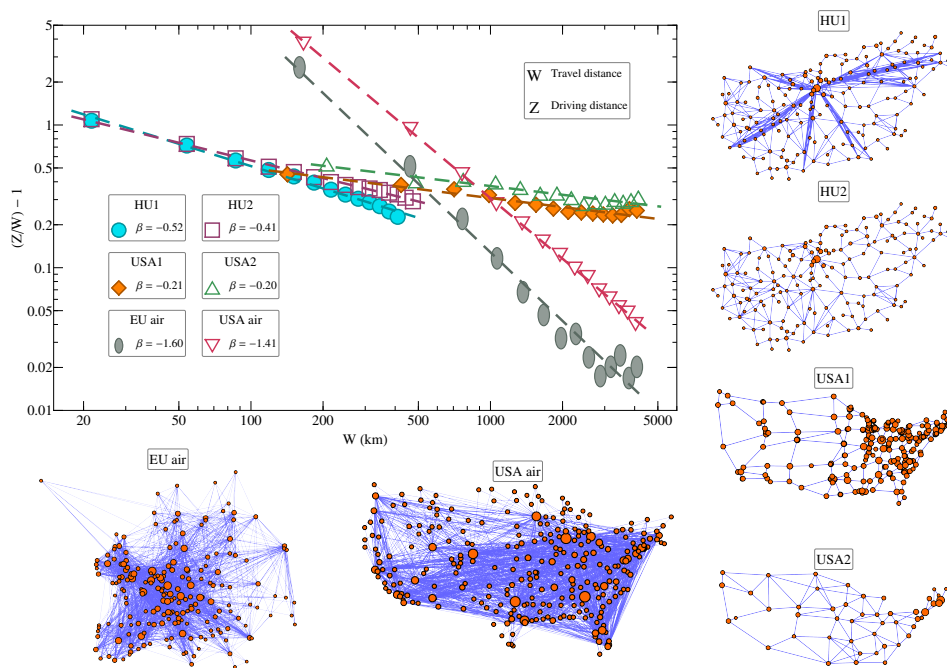


Figura 3.2: Topologia rețelelor de călătorie. Scalarea distanței de conducere (z) și distanței geodetică (w) pe diferite rețele de călătorie. Liniile indică o relație validă de $\frac{z}{w} = 1 + C \cdot w^{-\beta}$. Valorile de β sunt arătate în legendă. Arătăm aici și topologia rețelei a unor mobilități umane utilizate în acest studiu.

scalarea dintre distanță și mărimile vitezei (dreptunghiurile de pe figura 3.1). Observăm că tendința generală este o lege de putere cu un exponent mediu de 0,5 pentru întreaga scară a distanței și vitezei. Acest fenomen există și separat în diferite moduri de transport. Considerând drumurile cu rang similar (aceleași limite de viteză: HU2, USA1, USA2), se obțin tendințe de scalare cu exponenți nu foarte diferiți. În ceea ce privește tranzitul aerian, am utilizat doar informații de zboruri directe între aeroporturi, iar rezultatele arată o scalare cu exponent de aproximativ 0,25.

Topologia rețelelor pe care se desfășoară transportul este prima cauză evidentă a scalării dinamice netriviiale între distanță și timp. Nodurile nu sunt neapărat conectate pe drumuri drepte, de obicei, nu există rute directe între ele [30]. Pentru a călători de la un nod la altul, călătorii urmăresc o cale cu cea mai scurtă lungime dintre noduri: distanța de călătorie sau conducere este notată cu z . Pe măsură cum distanța geodetică a călătoriei w crește, distanța de conducere z se apropie mai mult în aproximația lui w . Evaluând topologia pentru anumite rețele de transport, observăm o relație de scalare destul de generală (figura 3.2):

$$\frac{z}{w} = 1 + C \cdot w^{-\beta} \quad (3.1)$$

Exponenții β corespunzători separă vizibil topologia rețelelor de tranzit rutier și aerian. Pentru călătoriile rutiere am observat $\beta \approx 0,2 - 0,3$, iar pentru călătorii aeriene $\beta \approx 1,4 - 1,5$. Valoarea z în cazul rețelelor de călătorie rutieră converge mai lent la w decât în cazul tranzitului aerian.

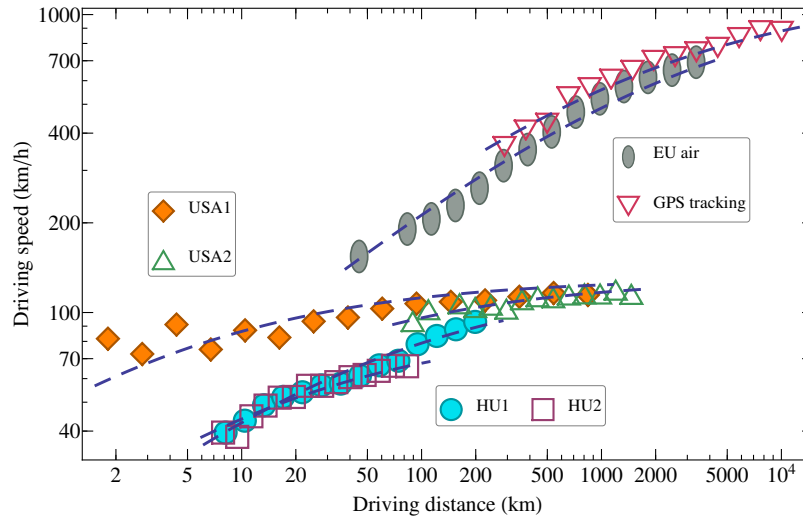


Figura 3.3: Viteza de călătorie. Viteza medie aparentă, u , în funcție de distanța de călătorie, z . Rezultatele sunt reprezentate pentru călătoria rutieră și tranzitul aerian împreună, ambele fiind bine descrise folosind formula din ecuația (3.2) cu $\alpha = 0.5$.

La călătoriile aeriene se observă creșterea vitezei aparente în funcție de distanță geodesică în ciuda faptului că doar zborurile directe (aproape în linie dreaptă) erau considerate între aeroporturi, pentru care se presupune că: $z = w$. Rezultatul este același, când luăm în considerare doar legăturile directe dintre nodurile pe rețea rutieră. În acest caz, viteza medie aparentă de călătorie pe o conexiune directă este definită ca $u = z/t$ și crește, de asemenea, odată cu distanța de transport, z . Figura 3.3 arată rezultatele în acest context.

Se poate argumenta că punctul de pornire și punctul de destinație provoacă întârzieri la fiecare conexiune. Mai mult, există, de asemenea, întârzieri în afara nodurilor pentru fiecare segment liniar. Acest lucru se datorează faptului că fluxul de trafic este diferit de ideal. În secții mai mari sunt așteptate întârzieri mai mari, dar creșterea nu este liniară. Presupunând că există o viteză de deplasare limitată într-un anumit segment, u_0 , și întârzierile cresc cu o tendință de tip lege de putere în funcție de lungimea segmentului, z : $t_{delay} = K \cdot z^\alpha$, găsim:

$$u = \frac{z}{\left(\frac{z}{u_0}\right) + t_{delay}} = \frac{1}{\left(\frac{1}{u_0}\right) + K \cdot z^{\alpha-1}} \quad (3.2)$$

Se obține o tendință de creștere când $\alpha < 1$. Pe figura 3.3 arătăm că ecuația (3.2) descrie bine toate datele experimentale de la traficul rutier până la traficul aerian. Pentru datele rutiere naționale (HU2) limita maximă era de $u_0 = 90$ km/h, în timp ce pentru datele de autostradă (HU1, USA1, USA2) am folosit $u_0 = 130$ km/h. Pentru călătoriile aeriene am utilizat $u_0 = 1200$ km/h ≈ 1 Mach. Un exponent $\alpha = 0.5$ oferă o descriere adecvată pentru toate datele, indicând o convergență cu limita de viteză u_0 .

În final, valoarea limitei de viteză crește, odată cu lungimea segmentului rutier sau al zborului. De

obicei pe segmentul rutier mai lung, limita de viteză este crescută: autostrăzile au segmente mai lungi și limite de viteză mai mari în comparație cu drumurile naționale. Pe segmente mai mari de călătorie aeriană, călătoresc avioanele mai rapide, un efect similar este valabil și pentru călătoriile cu trenul. Toate aceste efecte explică scalarea dinamică netrivială observată în mobilitățile umane.

3.2 Scalări în rețeaua Internetului, experiment și model

În secțiunea precedentă s-a discutat o scalare interesantă între timpul și distanța de călătorie, valabil pe 10 ordine de mărime în spațiu, pentru toate modurile de transport uman [9]. În continuare arătăm că o scalare similară poate fi găsită pentru dinamica datelor pe Internet [31]. Pentru a-l înțelege, propunem un model, care poate explica nu doar scalarea dinamică observată, dar și proprietățile topologice măsurate.

Experimentul nostru referitor la dinamica datelor pe Internet se bazează pe transferul și primirea pachetelor de mesaje ecou cu protocolul de control pe Internet (ICMP, Internet Control Message Protocol) [32]. Am utilizat comanda „ping” [33] pentru a măsura timpul de răspuns (timpul de călătorie dus-întors sau RTT a mesajelor ecou) la diferite servere de destinație. Cea mai frecventă unitate de măsurare a timpului este *ms*. În total 24700 de destinații erau alese din diverse locații de pe glob. Pozițiile lor globale (coordonatele GPS) erau determinate dintr-un tabel de adrese IP, folosind pagina web a IP2LOCATION [34]. Am determinat distanța geodetică d între routerele de sursă și destinație folosind aceste coordonate GPS. Comanda „traceroute” folosește același principiu de bază ca „ping”. Cererea de ecou ICMP primește RTT și în acest caz obținând hopuri intermediare, dezvăluind structura topologică la nivel de router a rețelei.

Studiile noastre se bazează atât pe experimentul „ping”, cât și pe rezultatele liber accesibile ale CAIDA UCSD IPv4 Routed / 24 Topology Set de date [35] obținute cu „traceroute”. Aceste studii arată că timpul mediu de călătorie (RTT) între routerele examinate crește cu o tendință subliniară în funcția distanței geodetice.

S-a aproximat trendul cu o regresie: $RTT = a \cdot d^{1/2}$. Coeficienții de determinare R^2 sunt: $R^2 = 0,98$ pentru datele de ping și $R^2 = 0,88$ pentru datele de traceroute (figura 3.4).

3.2.1 Modelul de Internet și rezultatele

Considerăm un model de rețea cu o regulă de conectare dinamică dintre noduri, pentru a explica această lege de scalare dinamică observată și celelalte proprietăți topologice a Internetului. În model, orașele sunt reprezentate de nodurile grafului. Conexiunile dintre ele corespund cablurilor de rețea. Nodurile N sunt răspândite uniform în spațiul Euclidian în această aproximare. Teritoriul considerat este un pătrat cu laturi de mărime unitară.

Folosim populația orașelor W_i , pentru a determina „raza de conectivitate”, ω_i , ca $\omega_i = \beta\sqrt{W_i}$.

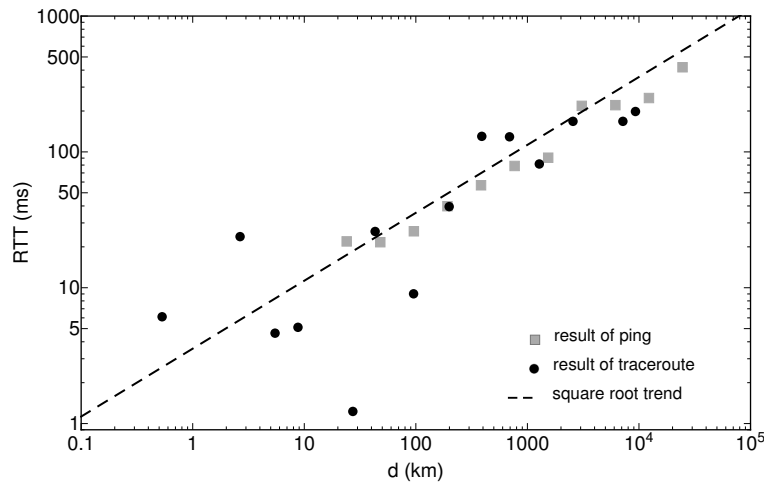


Figura 3.4: Date pentru transmisia pachetelor obținute cu instrucțiunile Ping și Traceroute. Timpul dus-întors la ambele experimente în funcție de distanță, folosind binarea logaritmică menționată în [9]. Linia punctată reprezintă tendința de lege de putere cu exponent $1/2$.

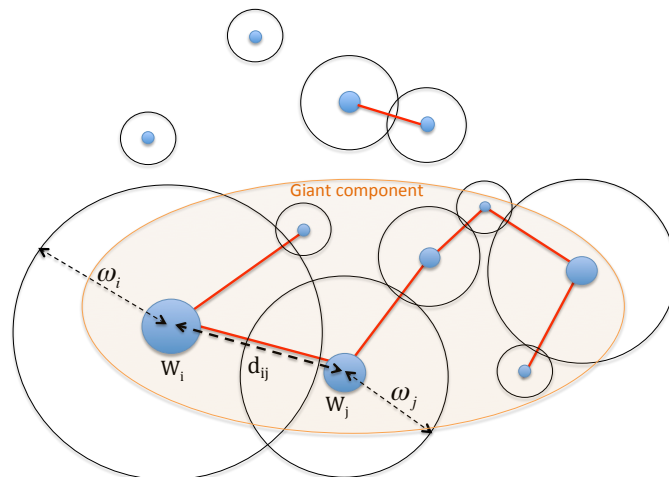


Figura 3.5: Schema de conectare în model. Aspecte principale și regula de conexiune a modelului.

W_i sunt valori alocate conform unei distribuții de tip Tsallis-Pareto cu exponentul $\alpha = 1$, adecvat pentru orașe mari după cum este argumentat în: [36, 37].

În mecanismul de conectare, pentru fiecare pereche de noduri calculăm:

$$f_{ij} = \frac{\omega_i + \omega_j}{d_{ij}} \quad (3.3)$$

Unde d_{ij} este distanța Euclidiană (sau lungimea liniei geodetice) între două orașe. Dacă $f_{ij} > 1$ legăm nodurile altfel ele rămân neconectate (figura 3.5).

Căutăm valoarea optimă β , care să fie capabilă să reproducă legile de scalare cunoscute și conexiunea medie a nodurilor (conexiuni/noduri) 8,68 observat în experimentele. Pentru fiecare grup de parametru am mediat rezultatele simulării pe 100 de realizări independente ale rețelei cu parametrii fixate de N și β . Pentru prelucrare se consideră doar componenta gigantică a grafului.

Pentru $N = 2400$ s-a ajuns la o concluzia, că proprietățile observate în experimente sunt cel mai bine reproduse de model cu pentru $\beta \approx 0.4$. Pentru rețelele construite cu $N = 8000$ și $\beta = 0.4$ componenții giganți au proprietăți statistic similare.

Distribuția conectivității nodurilor este o distribuție de tip Paretto-Tsallis (sau Lomax II):

$$p(k) = \frac{\alpha}{(\alpha - 1)\langle k \rangle} \left(1 + \frac{k}{(\alpha - 1)\langle k \rangle} \right)^{-1-\alpha} \quad (3.4)$$

Distribuția conectivității nodurilor obținute din rezultatele experimentale este reprezentată cu puncte negre din figura 3.6, și regresia Tsallis-Paretto (3.4) cu $\alpha = 1.23$ este indicat cu linie punctată roșie. Distribuțiile similare pentru conectivitatea nodurilor sunt de obicei acceptate în domeniul consacrat a rețelelor complexe [37, 38]. Distribuția numărului de conexiuni măsurate și cea oferită de model în comparație (figura 3.6) arată o concordanță acceptabilă, modelul redă astfel și proprietățile topologice găsite în [39].

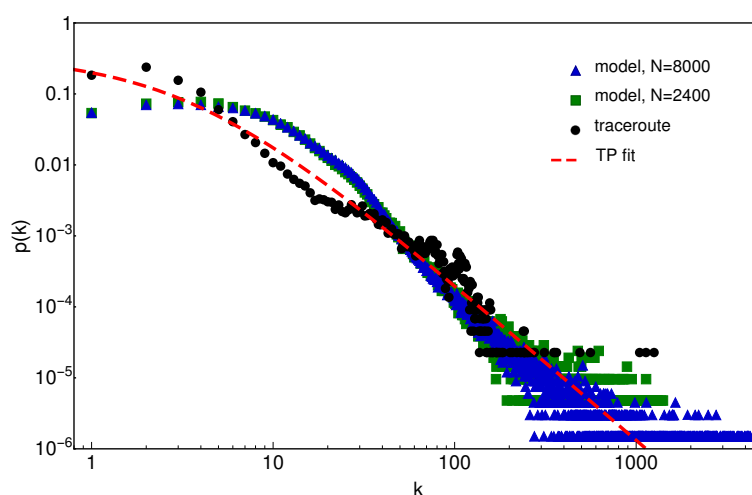


Figura 3.6: Distribuția conectivității nodurilor. Rezultatele din experimentul de traceroute marcat cu puncte negre și același distribuție produsă de model cu pătrate verzi pentru $N = 2400$ și triunghiuri albastre pentru $N = 8000$. Linia roșie punctată indică o regresie cu distribuția Tsallis-Paretto (3.4) cu $\alpha = 1.23$ și $\langle k \rangle = 8.68$. Coeficientul de determinare calculat pentru regresia Tsallis-Paretto și experiment este de $R^2 = 0.85$, iar pentru rezultatele modelului este totdeauna peste $R^2 > 0.9$.

Cea mai mare contribuție la timpul de răspuns este datorată perioadelor de așteptare suferite la routere. Astfel, putem presupune că timpul mediu de răspuns măsurat ar trebui să crească odată cu numărul de routere (hopurile) H întâlnite până la destinație. Într-adevăr, experimentele sugerează o relație de $\text{RTT} \propto H^\gamma$ (figura 3.7) cu $\gamma \approx 3/4$, cu un coeficient de determinare $R^2 = 0.98$. În consecință, de aici ar rezulta că numărul de hopuri vizitate crește în funcție de distanță cu un exponent de scalare: $(1/2)/(3/4) = 2/3$. Rezultatele din figura 3.8 validează această scalare.

Pentru determinarea celor mai scurte căi topologice între noduri, am utilizat metoda de căutare Breadth-first search implementată în pachetul Python Igraph [40]. Această este justificată de protocolul ICMP, vizitând cantitatea minimă de hopuri. RTT în funcție de distanță arată o tendință

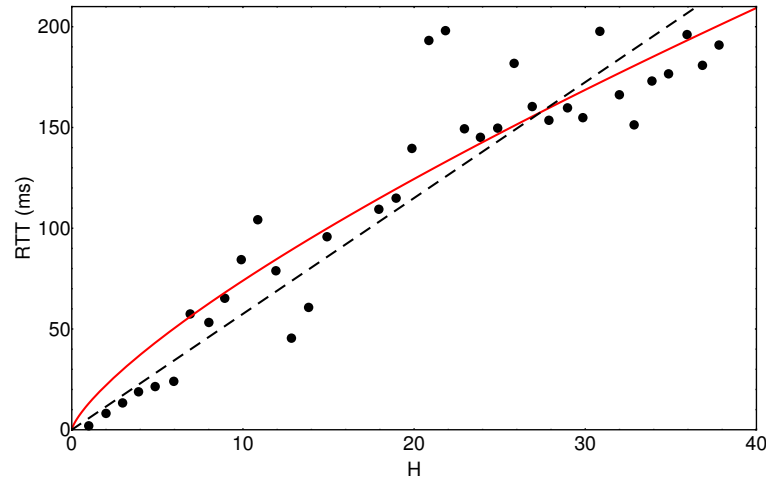


Figura 3.7: RTT mediu în funcție de numărul hopurilor stabilit în urma experimentului de tracero-ute. Linia punctată indică o proporționalitate liniară cu un coeficient de determinare de $R^2 = 0.96$, în timp ce linia roșie continuă indică o potrivire de tip lege de putere cu un exponent de $3/4$. Coeficientul de determinare a acestuia este de $R^2 = 0.98$.

de tip lege de putere, atât din rezultatele experimentale, cât și din cea obținută de modelul nostru. Această asemănare între experimentele și rezultatele modelului arată că un model simplu este capabil să capteze esența unei asemenea scalări.

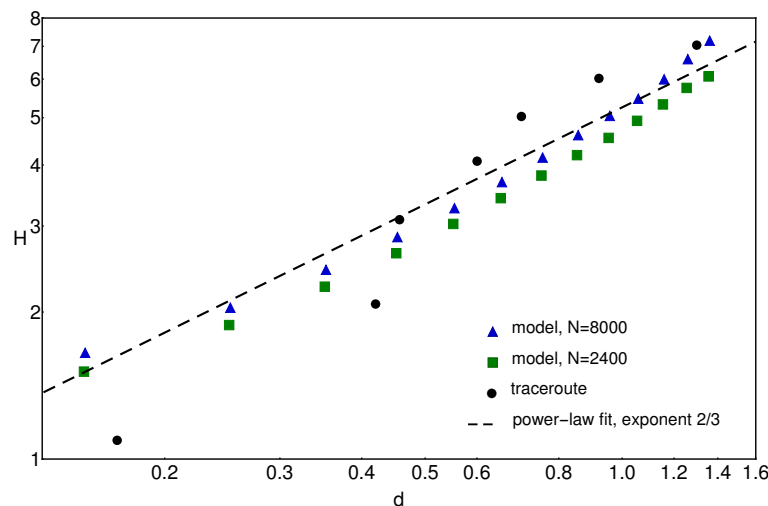


Figura 3.8: Numărul hopurilor în funcția de distanță. Punctele negre reprezintă datele obținute din experimentele de traceroute. Pătrate verzi ($N = 2400$) și triunghiuri albastre ($N = 8000$) reprezintă rezultatele modelului. Linia punctată sugerează o tendință a legii de putere având un exponent de $2/3$. Coeficientul de determinare a acestei regresii ($y = a \cdot x^{2/3}$) pentru datele de traceroute este $R^2 = 0.98$, în timp ce pentru rezultatele modelului sunt peste $R^2 = 0.99$. Distanțele geodetice prin experimentul traceroute sunt rescalate în pătratul de $(0, 1) \times (0, 1)$.

4

Concluzii

În cadrul tezei s-au studiat două problematici, ambele teme fiind legate de utilizarea metodelor fizicii statistice și computaționale în rețele complexe. Referitor la aceste două teme s-au ajuns la următoarele concluzii:

4.1 Detectarea comunităților în rețele complexe

S-a demonstrat că diagramele Voronoi pe grafuri sunt potrivite pentru detectarea comunităților în rețele complexe. Pentru a realiza acest lucru s-au identificat două cerințe: (1) definirea unui sistem metric de distanță adecvată între noduri; (2) identificarea nodurilor generatoare care indică „centrele” ale comunităților. Pentru măsurarea distanței, s-a ales inversul coeficientului de clusterizare a legăturilor, $1 / ECC$. Nodurile erau selectate ca generatoare atunci când densitățile lor locale relative erau maxime în vecinătatea lor pe o rază r . Acceptând acestea s-a arătat că tehnica noastră poate fi la fel de bună ca celelalte metode acceptate. Doar un singur algoritm a reușit să depășească performanțele metodei noastre. Partiționarea Voronoi prin creșterea razei r este cea mai bună strategie pentru această metodă de clusterizare.

S-a studiat teoretic și practic avantaje și dezavantaje al partiționării rețelelor cu diagrame Voronoi stohastice. În comparație cu alte măsuri de similitudine a nodurilor, [3, 41] coeziunea Voronoi obține informații locale din structura globală a rețelei. În secțiunea analitică a tezei s-a demonstrat partiționarea unui graf cu două module și caracteristicile introduse de conexiunile între aceste module. Rezultatele analitice erau confirmate cu simulări. În teză s-a prezentat o metodă care ajută la detectarea comunităților din informațiile matricei de coeziune.

4.2 Legi de scalare în rețele reale

S-a dovedit ipoteza conform căreia timpul călătoriilor la orice mijloc de mobilitate se scalează netrivial cu distanță. Viteza aparentă medie (raportul dintre distanța geodetică și durata deplasării pe rețea) crește monoton în forma unei legi de putere cu distanța. S-a demonstrat că atât structura topologică a rețelei rutiere, cât și factorii care influențează abaterea de la condiții ideale (efectele de întârziere ale nodurilor, limitele de viteză), contribuie la acest efect universal. Pentru a crește viteza aparentă, în primul rând, rețelele rutiere (sau conexiunile aeriene) trebuie optimizate astfel încât exponenții topologici ai rețelei β să fie maximizate. În acest caz, traseul de deplasare ar putea atinge traseul pe distanța geodetică, scăzând lungimea drumurilor. Planificarea geometriei rețelelor rutiere între orașele mari pare însă problematic, datorită distribuției lor spațiale fixe, care în mare parte predetermină structura interconectării lor. Cu toate acestea, la călătoriile aeriene, colaborările între diferite companii pot optimiza în acest sens.

Studiul experimental al dinamicii datelor pe Internet a condus la concluzii similare. Compararea unui model realist elaborat de noi cu rezultatele experimentale a adus la concluzia că modelul nostru simplu de cablare cu un parametru, încorporat în spațiul geometric este capabil să reproducă calitativ caracteristicile statistice descoperite în rețeaua de Internet construit pe nivelul routerilor. În acest sens, scalarea non-trivială a timpului mediu de răspuns a pachetului ecou în funcție de distanță geodetică se datorează topologiei specifice a rețelei. Diferența dintre exponenții de scalare referitoare la: tendința numărului de hop în funcție de distanță ($\approx 2/3$) și exponentul de scalare a timpului de răspuns în funcție de distanță ($\approx 1/2$), indică faptul că o întârziere medie constantă pe routere nu poate explica în totalitate scalarea observată. Pe lângă topologia rețelei trebuie să fie luat în considerare și alte aspecte pentru a construi un model mai realist. Acest lucru este oarecum similar cu concluziile de la investigarea scalărilor în mobilitățile umane.

5

Lista de publicații

Publicații relevante pentru teză

Lucrări științifice

- I. Papp, L. Varga, M. Afifi, I. Gere, and Z. Néda, Scaling in the space-time of the Internet. Scientific Reports, 9(1):9734, 2019. ISSN 2045-2322. (IF: 4.122)
- Z. I. Lázár, I. Papp, L. Varga, F. Járαι-Szabó, D. Deritei, M. Ercsey-Ravasz, Stochastic graph Voronoi tessellation reveals community structure, Phys. Rev. E, 95, 2017 (IF: 2.353)
- L. Varga, A. Kovács, G. Tóth, I. Papp, Z. Néda, Further We Travel the Faster We Go, PLoS ONE 11(2): e0148913, 2016 (IF: 2.776)
- D. Deritei, Z.I. Lázár, I. Papp, F. Jarai-Szabo, R. Sumi , L. Varga, E. Regan, M. Ercsey-Ravasz, Community detection by graph Voronoi diagrams, New Journal of Physics, 2014 (IF: 3.773)

Prezentări la conferințe

- I. Papp, M. Afifi, L. Varga, I. Gere and Z. Néda, Nontrivial dynamical scaling in the Internet: experiments and a simple model, talk, 12th Joint Conference on Mathematics and Computer Science Cluj-Napoca, June 14 – 17, 2018.
- I. Papp, M. Afifi, L. Varga, I. Gere, Z. Néda, Scaling in the space-time of the Internet, Poster 55, MECO43, 2018 May

-
- Varga L, Kovács A, Tóth G, Papp I, Néda Z, Velocity versus distance scaling in human travel, poster, XXXVI Dynamics Days Europe, Corfu, Greece, (2016)
 - Z. I. Lázár, I. Papp, L. Varga, F. Járari-Szabó, D. Deritei, M. Ercsey-Ravasz, Community detection using stochastic graph Voronoi tessellation, poster, NetSci, (2015)
 - D. Deritei, Z. I. Lázár, I. Papp, F. Jarai-Szabo, R. Sumi, M. Ercsey-Ravasz; Community detection by graph Voronoi diagrams; NetSci 2013, International School and Conference on Network Science, June 3-7 2013, Copenhagen, Denmark

Alte publicații din programul de studii doctorale

Lucrări științifice

- L.P Csernai, M. Csete, I.N. Mishustin, A. Motornenko, I. Papp, L.M. Starov, H. Stöcker, N. Kroó, Radiation dominated implosion with flat target, Physics of Wave Phenomena, February 2020, accepted for publication (IF:0.641)
- L.P Csernai, N. Kroó and I. Papp, Radiation dominated implosion with nano-plasmonics, Laser and Particle Beams, Volume 36, Issue 2, June 2018 , pp. 171-178 (IF: 1.194)
- K.Z. Rónai, A.Szentkirályi A. S. Lázár Z. I. Lázár I. Papp, F. Gombos, R. Zoller, M. E. Czira, A. V. Lindner, I. Mucsi, R. Bodizs, M. Z. Molnar, M. Novák, Association of symptoms of insomnia and sleep parameters among kidney transplant recipients, Journal of Psychosomatic Research, 95-104, 2017 (IF: 2.722)
- L.P. Csernai, I. Papp, S.F. Spinnangr and Y. Xie, Physical Basis of Sustainable Development, Int. J. of Central European Green Innovation 42, 2016, 39-50.

Prezentări la conferințe

- L.P. Csernai, N. Kroó and I. Papp, Applications of Advances in Relativistic Fluid Dynamics to Laser Fusion, 6th International Conference on New Frontiers in Physics, Kolymbari, Crete (ICNFP 2017)
- D. Deritei, L. Varga, I. Papp, F. Járari-Szabó, Z. I. Lázár, R. V. Florian and M. Ercsey-Ravasz, Genetic-like algorithm applied on citation networks for evaluating scientific publications, poster, Conference on Complex Systems, Amsterdam, Netherlands, (2016)

-
- I. Papp and L.P. Csernai, Inertial Confinement Fusion: Radiation Dominated Implosion, talk, Exploring the Shores of Fundamental Matter: Advances around the Northern Seas (NorSAC-2015), Bergen, Jul. 31, 2015
 - L. Varga, F. Jarai-Szabo, D. Deritei, Z. I. Lázár, I. Papp, R. Florian, M. Ercsey-Ravasz , Local Cluster Detection Method for Normalizing Scientometric Indicators, poster, NetSci, (2015)
 - I. Papp, M. Ercsey-Ravasz, D. Deritei, R. Sumi, F. Jarai-Szabo, R. V. Florian, A. I. Cabuz, Z. I. Lázár; The PH-Index: Hirsch Index of Individual Publications, poster; ISSI 2013, 14th International Society of Scientometrics and Infometrics Conference, 15-19 July 2013, Vienna, Austria

Brevete

- L. P. Csernai, N. Kroo and I. Papp, Procedure to improve the stability and efficiency of laser-fusion by nano-plasmonics method, Patent No. P1700278/3 of the Hungarian Intellectual Property Office, 2017.

Bibliografie

- [1] Newman, MEJ. Networks: An Introduction. Oxford University Press, 2010.
- [2] Barabasi, AL. Network science. 2012.
- [3] Santo Fortunato. Community detection in graphs. Physics Reports, 486(3):75 – 174, 2010. ISSN 0370-1573.
- [4] Han, J. Kamber, M. Pei, J. Data mining: Concepts and techniques. 2011.
- [5] Theodoridis, S. Koutroumbas, K. Pattern recognition. 2008.
- [6] Aurenhammer, F. Voronoi diagrams - a survey of a fundamental geometric data structure. ACM Computing Surveys, 23(3):345–405, 1991.
- [7] Kung, K.S., Greco, K., Sobolevsky, S., Ratti, C. Exploring universal patterns in human home-work commuting from mobile phone data. PLoS ONE, 9(6), 2014.
- [8] Gallotti, R. & Barthélemy, M. Anatomy and efficiency of urban multimodal mobility. Sci. Rep., 4:6911, 2014.
- [9] Levente Varga, András Kovács, Géza Tóth, István Papp, and Zoltán Néda. Further we travel the faster we go. PLOS ONE, 11(2):1–9, 02 2016.
- [10] Barabasi AL and Albert R. Emergence of scaling in random networks. Science, 286:509–512, 1999.
- [11] SH. Yook, H. Jeong, and AL. Barabasi. Modeling the Internet’s large-scale topology. PNAS, 99:13382–13386, 2002.
- [12] F. Castellano, C. Cecconi, F. Loreto, V. Parisi, D. Defining and identifying communities in networks. Proc. Natl. Acad. Sci. USA, 101:2658, 2004.
- [13] Newman, MEJ. Modularity and community structure in networks. PNAS, 103(23):8577–8582, 2006.
- [14] Zachary WW. An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33:452–473, 1977.
- [15] White, JG. Southgate, E. Thompson, JN. and Brenner, S. . The structure of the nervous system of the nematode caenorhabditis elegans. Phil. Trans. R. Soc. London, 314, 1986.

-
- [16] Watts, DJ and Strogatz, SH. Collective dynamics of 'small-world' networks. Nature, 393: 440–442, 1998.
- [17] Jeong, H. Mason, S. Barabási, AL and Oltvai, ZN. Lethality and centrality in protein networks. Nature, 411:41–42, 2001.
- [18] Newman, MEJ. The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. USA, 98:404–409, 2001.
- [19] Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 u.s. election: Divided they blog. In Proceedings of the 3rd International Workshop on Link Discovery, LinkKDD '05, pages 36–43, New York, NY, USA, 2005. ACM. ISBN 1-59593-215-1.
- [20] Blondel, VD. Guillaume, JL. Lambiotte, R. Lefebvre, E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), 2008.
- [21] Raghavan, UN. Albert, R. Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. Phys. Rev. E, 76:036106, 2007.
- [22] Jierui Xie and Boleslaw K. Szymanski. Towards linear time overlapping community detection in social networks. In Proceedings of the 16th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining - Volume Part II, PAKDD'12, pages 25–36, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-30219-0.
- [23] J. Xie, B. K. Szymanski, and X. Liu. Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In 2011 IEEE 11th International Conference on Data Mining Workshops, pages 344–349, Dec 2011.
- [24] Ahn, YY. Bagrow, JP and Lehmann, S. Link communities reveal multiscale complexity in networks. Nature, 466:761–764, 2010.
- [25] Rosvall, M and Bergstrom, CT. Maps of random walks on complex networks reveal community structure. PNAS, 105:1118–1123, 2008.
- [26] Rosvall, M. Axelsson, D. and Bergstrom, CT. The map equation. Eur. Phys. J. Special Topics, 178:13–23, 2009.
- [27] Zsolt I. Lázár, István Papp, Levente Varga, Ferenc Járai-Szabó, Dávid Deritei, and Mária Ercsey-Ravasz. Stochastic graph Voronoi tessellation reveals community structure. Phys. Rev. E, 95:022306, Feb 2017.
- [28] Lancichinetti, A. Fortunato, S. Radicchi F. Benchmarks graphs for testing community detection algorithms. Phys. Rev. E, 78, 2008.
- [29] Lancichinetti, A. Fortunato, S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. Phys. Rev. E, 80, 2009.

-
- [30] Marc Barthelemy. Spatial networks. Physics Reports, 499(1):1 – 101, 2011. ISSN 0370-1573.
- [31] István Papp, Levente Varga, Mounir Afifi, István Gere, and Zoltán Nédá. Scaling in the space-time of the Internet. Scientific Reports, 9(1):9734, 2019. ISSN 2045-2322.
- [32] J. Postel. Internet Control Message Protocol. RFC, 777:1–14, 1981.
- [33] Linux man page for ping command. <https://linux.die.net/man/8/ping>.
- [34] The ip adress databases. <http://www.ip2location.com>.
- [35] The CAIDA UCSD IPv4 Routed /24 Topology Dataset - 2017. http://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml.
- [36] M.E. Moses E.H. Decker, A.J. Kerkhoff. Global patterns of city size distributions and their fundamental drivers. Plos One, 22:e934, 2007.
- [37] T-S. Biro and Z. Neda. Unidirectional random growth with resetting. Physica A, 499:335–361, 2018.
- [38] Barabasi AL and Albert R. Emergence of scaling in random networks. Science, 286:509–512, 1999.
- [39] Mahadevan, P. and Krioukov, D. and Fomenkov, M. and Huffaker, B. and Dimitropoulos, X. and Claffy, K. and Vahdat, A. The Internet as-level topology: Three data sources and one definitive metric. ACM SIGCOMM Computer Communication Review (CCR), 36(1):17–26, 2006.
- [40] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. InterJournal, Complex Systems, 1695(5):1–9, 2006.
- [41] M. E. J. Newman. Detecting community structure in networks. The European Physical Journal B, 38(2):321–330, Mar 2004.