
COMMUNITY DISTRIBUTION ON WEIGHTED SOCIAL NETWORKS AND TIME RESOLUTION

Klotilda Nikaj

University of Tirana, Albania, klotildanikaj@hotmail.com

Margarita Ifti

University of Tirana, Albania, margarita.iftit@rambler.ru

Abstract: The main aim of this paper is improving the efficiency and accuracy of community detection in complex networks. A popular method for detecting communities consists of maximizing a quality function known as modularity. We proposed a new algorithm, which is based on the idea that communities could be detected from subnetworks by comparing the internal and external density of each subnetwork. In our method, similar nodes are firstly gathered into meta-communities, which are then decided to be merged through a label propagation process, until all of them meet our community structure. An important meso-scale feature of these networks is measured through their community structure, which defines groups of strongly connected nodes that exist within and across network. Because subnetwork edges can describe relationships between different modalities, scales, or time points, it is essential to understand how communities change and evolve across them. Here, we expose an upper bound for time resolution beyond which community changes across layers cannot be detected. This upper bound has non-trivial, purely multilayer effects and acts as a resolution limit for detecting evolving communities. Our findings not only represent new theoretical considerations but also have important practical implications for choosing multilayer networks to model real-world systems whose communities change across time or modality. Our algorithm requires neither any priori information of communities nor optimization of any objective function. Experimental results show that, our algorithm performs quite well and runs extremely fast, compared with several other popular algorithms. By tuning time as a resolution parameter, we can also observe communities at different scales, so this could reveal the hierarchical structure of the network

Keywords: community structure, time resolution, weighted networks

SHPËRNDARJA E KOMUNITETEVE NË RRJETET SOCIALE TË PESHUARA DHE KOHA SI PARAMETER REZOLUCIONI

Klotilda Nikaj

Universiteti i Tiranës, Albaniaklotildanikaj@hotmail.com

Margarita Ifti

Universiteti i Tiranës, Shqipëri, margarita.iftit@rambler.ru

Përmbledhje: Qëllimi kryesor i këtij punimi lidhet me përmirësimin e dedektimit të komuniteteve në rrjetet sociale komplekse, duke përmirësuar një algoritëm që tashmë është përdorur edhe më përpara në kërkime të ngjashme i cili njihet ndryshe si funksioni i modularitetit. Modulariteti është një funksion i njohur që përdoret për të dedektuar komunitetet. Algoritmi i propozuar prej nesh bazohet në idenë se komunitetet mund të dedektohen duke krahasuar dendësinë e nyjeve dhe lidhjeve të nën-rrjeteve që e përbëjnë. Bazuar në këtë metode, nyjet e ngjashme grupohen fillimisht me njëra-tjetrën në meta-komunitete, të cilat në një hap të mëvonshëm bashkohen me njëra-tjetrën derisa të shfaqet struktura e komunitetit e kërkuar. Duke qënë se lidhjet ndërmjet nyjeve të rrjetës mund të jenë nga më të ndryshmet, edhe dinamika e tyre në kohë do ndryshojë dukshëm. Algoritmi që kemi modifikuar nuk kërkon informacione paraprake të komuniteteve as optimizimin e funksioneve, mjafton të dimë dendësinë e lidhjeve të nyjeve. Rezultatet eksperimentale tregojnë se ky algoritëm përformon shumë më shpejt krahasuar me algoritmet e tjera të njohura. Rezolucioni limit i modularitetit, vendos një vlerë limit për përmasat e komunitetit më të vogël që mund të veçojmë nga rrjeti me anë të optimizimit të funksionit të modularitetit. Duke përdorur kohën si parametër rezolucioni, algoritmi na jep mundësinë të vëzhgojmë strukturën e komunitetit në rreth të ndryshme hierarkike.

Fjalë kyçe: struktura e komunitetit, rezolucioni kohor, rrjete të peshuara

1. HYRJE

Në dekadat e fundit, janë propozuar shumë metoda për të dedektuar strukturën e komunitetit si për shembull agregimi spektral i nyjeve, dinamika apo metodat e bazuara në topologjinë e rrjetës. Në këto grupe metodash të propozuara nga shkencëtarë të ndryshëm gjejmë edhe përafrimin që i bëhet funksionit të modularitetit për të

dedektuar strukturën e komunitetit, funksion i cili e shfaq problemin kryesor tek rezolucioni[1,2,3]. Rezolucioni limit i modularitetit, vendos një vlerë limit për përmasat e komunitetit më të vogël që mund të veçojmë nga rrjeti me anë të optimizimit të funksionit. Kjo vlerë limite vendoset nga faktori $1/2m$ tek P_{ij} , e cila implikon se modulariteti varet nga përmasat e rrjetit, dhe jo vetëm nga karakteristikat lokale të agjentëve.

Aftësia e rrjetit për të ruajtur karakteristikat e tij, pasi fshihet një fraksion nyjesh, zakonisht njihet me termin *Qëndrueshmeri* (robustness). Kjo veçori e rrjetit është ndër të parat të eksploruara në literaturën[5,6] e rrjeteve komplekse meqenëse lidhet drejtpërdrejtë me funksionimin e duhur të rrjetës dhe mund të gjendet në versione të ndryshme:

Qëndrueshmeria statike, fshirja e nyjeve pa qënë nevoja të rishpërndahet ndonjë parametër i rrjetës.

Qëndrueshmeria dinamike, në këtë rast merret parasysh dinamika e parametrave të rrjetës, p.sh rrjedhja ose dinamika e informacionit në rrjetë.

Trajtimi teorik i qëndrueshmerisë të rrjeteve sociale mund të përafrohet duke përdorur teorinë e fushës mesatare. Formulimi i fraksionit kritik të nyjeve që duhen fshirë, në mënyre të tillë që rrjeti t'i eliminohet mundësia për të shkuar nga një nyje në një tjetër duke ndjekur një lidhje të caktuar e ka origjinën në teorinë e grafeve në vitin 1995, edhe pse aplikimi i parë praktik u bë në rrjetin e internetit[7,8,9]. Fraksioni kritik i nyjeve rastësore që duhen fshirë në rrjetë jepet me ekuacionin:

$$f_k = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$$

Ku $\langle k \rangle$ është mean degree, ndërsa $\langle k^2 \rangle$ katrori i shkallës mesatare e cila përkufizohet në teorinë e grafeve (Newman 2003):

$$\langle k^2 \rangle = \frac{\sum_{i=1}^N k_i^2}{N}$$

ku N është numri i nyjeve të rrjetit. Ky ekuacion tregon numrin e nyjeve që duhet të fshihen rastësisht nga rrjeti në mënyre që të pengohet kalimi nga një nyje në një tjetër duke ndjekur një kalim rastësor.

2. MATERIALET DHE METODAT

Me përkufizim *komunitet* quhet një nëngraf i rrjetës, në të cilin densiteti i brendshëm i lidhjeve të rrjetës është dukshëm më i lartë se sa densitetit i jashtëm. Në trajtë të përgjithshme ky përkufizim përshkruhet nga ekuacioni i mëposhtëm i cili njohet ndryshe si funksioni i modularitetit:

$Q = \text{densiteti i lidhjeve brenda komunitetit} - \text{densiteti i lidhjeve në një rrjetë rastësore}$

$$Q = 1/2m \sum_c \sum_{i,j} [A_{ij} - P_{ij}]$$

Zakonisht zgjidhet funksioni $P_{ij} = \frac{k_i k_j}{2m}$ që të arrijë të marrë parasysh heterogjenitetin e lidhjeve të rrjetës, ndërsa

$$A_{ij} = k_i \text{ është fuqia e nyjes } i.$$

Supozojmë se kemi procese identike homogjene Poissoni në secilën nyje të rrjetit, të tilla që një agjent kërcen me një probabilitet konstant nga secila nyje, duke i korresponduar proceseve continuous-time që i binden ekuacionit Kolmogorov:

$$\dot{p}_i = \sum_j \frac{A_{ij}}{k_j} p_j - p_i$$

Zgjidhja stacionare për këto procese vazhdon të jetë $p_i^* = k_i / 2m$. Përkufizojmë probabilitetin që një agjent të mbetet brenda një komuniteti gjatë intervalit kohor t , duke mos marrë parasysh probabilitetin që një ngjarje e tillë të ndodhë rastësisht në kushte stacionare:

$$\sum_c \sum_{i,j} [(e^{t(B-I)}) \frac{k_j}{2m} - \frac{k_i}{2m} \frac{k_j}{2m}]$$

Ku $B = \frac{A_{ij}}{k_j}$ dhe I është matrica njësi. Ky funksion është jo-rritës dhe shkon drejt zeros kur $t \rightarrow \infty$, me fjalë të tjera kushtet fillestare humbasin kur variojmë kohën në pafundësi.

Shohim rastin limit kur $t = 0$, funksioni i mësipërm shkruhet:

$$1 - \sum_c \sum_{i,j} \frac{k_i}{2m} \frac{k_j}{2m}$$

Ku shikojmë se ky funksion maskimizohet kur secila nyje ndodhet në komunitetin e vet përkatës në çastin $t = 0$. Nga ana tjetër, në rastin kur $t \rightarrow \infty$, numri i komuniteteve zvogëlohet me rritjen e kohës, nga një funksion me N -komunitete me një nyje në $t = 0$ me të pakten dy komunitete kur $t \rightarrow \infty$. Në këtë kuptim, koha mund të shikohet si një parameter intrisik rezolucioni, sipas të të cilit shfaqen struktura hierarkike në rrjetë, ku nëse shfaqet një strukture e vetme, rezultati konsiderohet i pakënaqshem.

Koha në secilin hap të simulimit ndjek rregullat e mëposhtme:

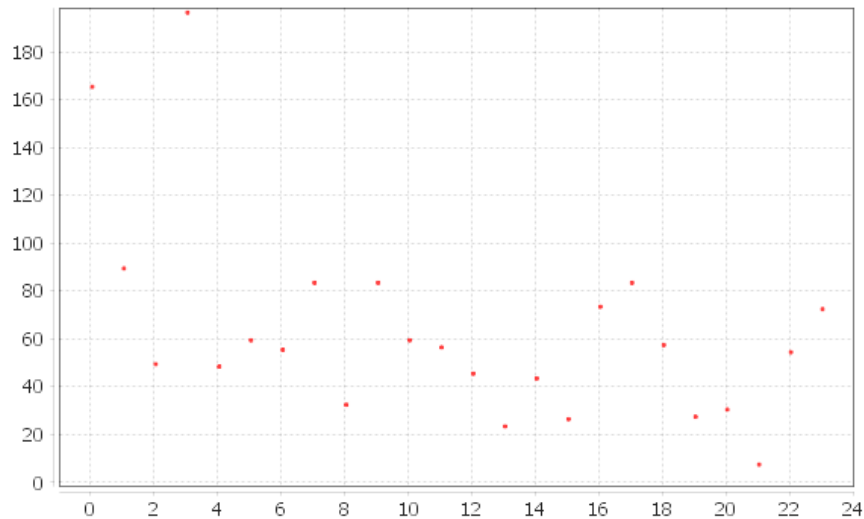
1. Llogaritja e ngjashmërisë së nyjeve merr kohë e rendit $O(km)$. Për secili set nyjesh, ne duhet të llogarisim ngjashmërinë së dendësisë me nyjet fqinjë, dhe kufiri i sipërm i kohës së kompleksikologjisë është i rendit $O(k)$.
2. Agregimi i rrjetës duhet të bëhet duke marrë parasysh të gjitha lidhjet e nyjeve të rrjetës, dhe koha që duhet për të përfunduar ka rendin $O(m)$.
3. Agregimi i rrjetës ndodh kur për secilën nyje kemi më së shumti k -iteracione, ku k - është numri i fqinjëve, nga ku rrjedh se kufiri i sipërm i kohës së kompleksikologjisë është i rendit $O(kn)$.

Hapat 2 dhe 3 përsëritën aq here sa të arrihet kriteri i vendosur gjatë simulimit i përfutimit të komuniteteve. Zakonisht ky algoritëm konvergjon pas një numri të vogël iteracionesh.

3. REZULTATET

Përdoruesit e rrjetave sociale bashkohen me grupe që kanë interesa të përbashkëta duke u aktivizuar me anë të komenteve apo pëlqimeve që ju japin postimeve të ndryshme. Në këtë punim ne kemi marrë në shqyrtim një grup të dhënash nga një database i hapur në Universitetin e Stanfordit[4]. Të gjitha të dhënat janë të anonamizuara, do të thotë që ne mund të identifikojmë nqs dy përdorues bëjnë pjesë në të njëjtin komunitet (psh jan pjesë e së njëjtës parti politike, por nuk cilësohet se cila parti politike). Të dhënat përbëhen nga 1034 nyje dhe 53498 lidhje.

Fig.1 Përdorim lidhjet e peshuara të rrjetit me rezolucion 0.5, numri i komuniteteve rezulton 24 (boshti y nr i nyjeve në një komunitet, boshti x-numri i komuniteteve)



Me rritjen e rezolucionit shikojmë se si ndryshon numri i komuniteteve në grafiket më poshtë:

Fig. 2 Rezolucioni 0.8, nr i komuniteteve 15 (boshti y nr i nyjeve për komunitet, boshti x-nr i komuniteteve)

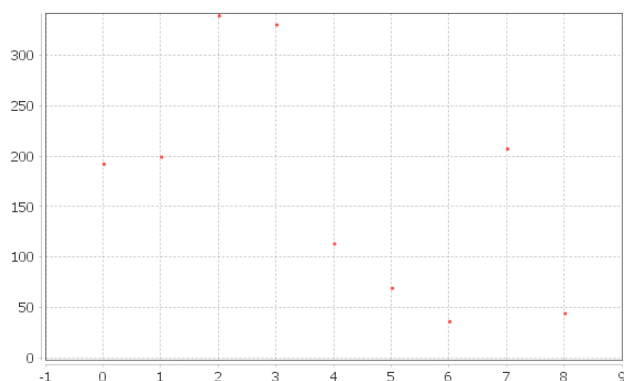


Fig. 3 Rezolucioni 1, nr i komuniteteve 13 (boshti y nr i nyjeve për komunitet, boshti x-nr i komuniteteve)

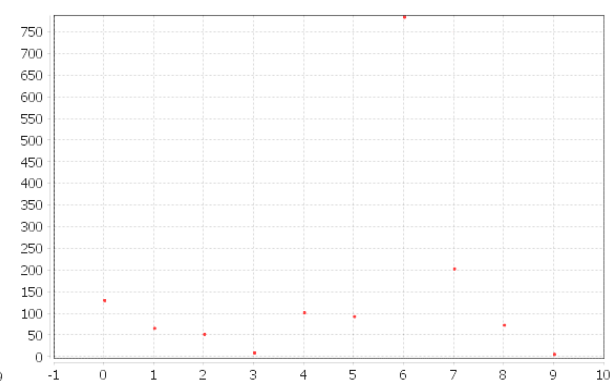


Fig. 4 Rezolucioni 1.2, nr i komuniteteve 9 (boshti y nr i nyjeve për komunitet, boshti x-nr i komuniteteve)

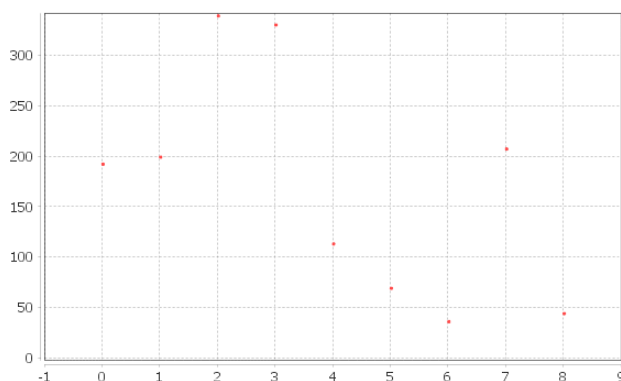
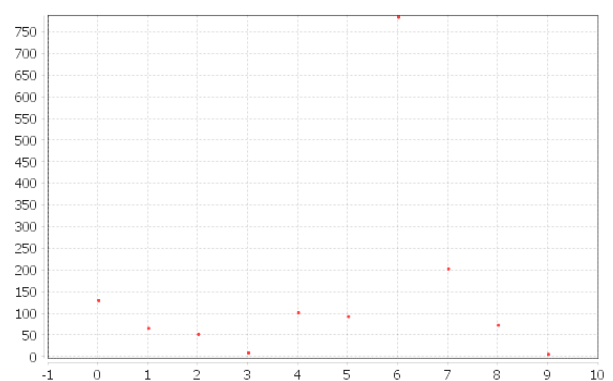


Fig. 5 Rezolucioni 1.5, nr i komuniteteve 10 (boshti y nr i nyjeve për komunitet, boshti x-nr i komuniteteve)



Në grafikun e mëposhtem, shikojmë se si ndryshon numri i komuniteteve të rrjetit, në rastin kur rezolucioni ka vlerën 1 por vendosim të fshijmë tre nga nyjet me numrin më të lartë të lidhjeve (Qëndrueshmeria statike)

Fig. 6 Pasi kemi fshirë 3 nyjet kryesore të rrjetit, shikojmë se si rrjeti copëzohet në afërsisht 130 komunitete (boshti y nr i nyjeve për komunitet, boshti x-nr i komuniteteve)



Në figurën e mësipërme flasim për fshirje të nyjeve jo rastësore, ose e thënë ndryshe merret në konsiderate sulmi që mund t'i bëhet rrjetës. Toleranca ndaj sulmeve mund të përshkruhet si aftësia e rrjetit për të ruajtur karakteristikat e tij kur bëhet një fshirje e qëllimshme e nyjeve.

Tabela 1

Numri i komuniteteve	Modulariteti	Rezolucioni	Modulariteti me rezolucion
24	0.451	0.5	0.195
15	0.464	0.8	0.352
13	0.450	1.0	0.450
9	0.450	1.2	0.576
10	0.397	1.5	0.782
138	0.567	1.0	0.567

Në tabelën 1 japim në mënyrë të përmbledhur rezultatet e nxjerra përsa i përket numrit të komuniteteve dhe varësinë që ka shfaqja e strukturës së komunitetit në një rrjetë sociale nga funksioni i modularitetit si edhe nga koha si parametër rezolucioni i këtij funksioni.

4. DISKUTIME

Nyjet e lidhura tërheqin njëra tjetrën dhe shtojnë nyjet e palidhura duke formuar grupe të cilat karakterizohen nga nyje që kanë një numër më të madh lidhjesh nqs i përkasin një komuniteti se ato që nuk bëjnë pjesë në komunitet. Pra, komunitetet shfaqen si grupe nyjesh, që karakterizohen nga një densitet i caktuar. Për rrjetet sintetike, kemi testuar modelin e propozuar nga Girvan dhe Newman me strukture komuniteti heterogjene. Për të testuar qendrueshmerinë e strukturës së komunitetit, kemi fshirë nyjet me numrin më të madh të lidhjeve në rrjet, dhe shikojmë se si nga një rrjete me 9 komunitete fillestare(shiko fig 4), shpërbashkohet në një rrjetë me 138 komunitete(shiko fig. 6). Në figuren 6 flasim për fshirje të nyjeve jo rastësore, ose e thënë ndryshe merret në konsiderate sulmi që mund t'i bëhet rrjetës. Toleranca ndaj sulmeve mund të përshkruhet si aftësia e rrjetit për të ruajtur karakteristikat e tij kur bëhet një fshirje e qëllimshme e nyjeve. Në tabelën 1 shohim në trajtë të përmbledhur si ndikon koha si parametër rezolucioni në numrin e komuniteteve, si edhe në funksionin e modularitetit. Gjithashtu, duke vëzhguar grafiket, mund të vërejmë se duke përdorur kohën si parametër rezolucioni, agregimi i nyjeve në strukturën e komunitetit bëhet gjithnjë e më i dukshëm duke rritur performancën e funksionit të modularitetit.

5. PËRFUNDIMET

Koha si një parameter rezolucioni vendos një vlerë limit për përmasat e komunitetit më të vogël që mund të veçojmë nga rrjeti me anë të optimizimit të funksionit të modularitetit. Duke përdorur kohën si parametër rezolucioni, algoritmi na jep mundësinë të vëzhgojmë strukturën e komunitetit në rreth të ndryshme hierarkike.

REFERENCAT

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* (2008). P10008.
- Han, J. (2016). Multi-resolution community detection in massive networks. *Sci. Rep.* 6, 38998; doi: 10.1038/srep38998.
- Heeren, A., & McNally, R. (2018). Social Anxiety Disorder as a Densely Interconnected Network of Fear and Avoidance for Social Situations
- Liu, W., Jiang, X., Pellegrini, M., & Wang, X. (2016). Discovering communities in complex networks by edge label propagation. *Scientific Reports* 6, 22470 EP–.
- Leskovec, J., & Krevl, A. (2014). SNAP Datasets: Stanford large network dataset collection <http://snap.stanford.edu/data> .
- Luo, R., Nettasinghe, B., & Krishnamurthy, V. (2021). Controlling Segregation in Social Network Dynamics as an Edge Formation Game
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* 69, 026113.
- Wenlin, L. (2017). Social Network theory
- Xiang, J., Li, H., & Li, J-M. (2018). [Critical analysis of \(Quasi-\)Surprise for community detection in complex networks](#).