

Bài thực hành số 2:

Phân tích cộng đồng trong mạng xã hội sử dụng Gephi

Phần 1: Phân tích cấu trúc mạng Tính toán và so sánh các độ đo tính trung tâm

(Centrality Measures) sau:

- Degree Centrality
 - Betweenness Centrality
 - Closeness Centrality
- Hãy xác định 3 nút có độ trung tâm cao nhất theo mỗi độ đo và giải thích ý nghĩa của chúng trong mạng lưới.

Dữ liệu sử dụng là **Political Blogs Network**.

1. Degree Centrality

Id	Label	Interval	value	source	Degree ▾
855	blogsforbush.com		1	BlogPulse,Campaig..	467
155	dailykos.com		0	LeftyDirectory,Lab...	383
1051	instapundit.com		1	BlogPulse	362
55	atrios.blogspot.co...		0	BlogPulse,LeftyDir...	350
641	talkingpointsmem...		0	BlogPulse,LeftyDir...	282
729	washingtonmonthl...		0	BlogPulse,LeftyDir...	256
963	drudgereport.com		1	LabeledManually	243
1245	powerlineblog.com		1	BlogPulse	235
1153	michellemalkin.co...		1	BlogPulse	228
1041	hughhewitt.com		1	Blogarama,Labele...	225

Degree Centrality đo số lượng kết nối trực tiếp mà một nút có với các nút khác. Những nút có Degree cao thường là trung tâm giao tiếp, nhận được nhiều sự chú ý hoặc tương tác từ các nút khác.

- blogsforbush.com: Có số lượng kết nối lớn nhất, thể hiện rằng nó là một blog rất phổ biến hoặc có vai trò quan trọng trong việc lan truyền thông tin trong mạng.

- [dailykos.com](#): Được kết nối với rất nhiều blog khác, đại diện cho một blog có lượng truy cập cao hoặc trung tâm trong cộng đồng chính trị trực tuyến.
- [instapundit.com](#): Cũng là một blog quan trọng với số lượng kết nối lớn, có thể đóng vai trò trong việc kết nối các nhóm khác nhau.

2. Betweenness Centrality

Id	Label	Interval	value	source	Betweenness Centrality ▾
855	blogsforbush.com	1		BlogPulse,Campaig..	218464.048305
55	atrios.blogspot.co...	0		BlogPulse,LeftyDir...	90985.835827
1051	instapundit.com	1		BlogPulse	76270.025259
155	dailykos.com	0		LeftyDirectory,Lab...	54982.016242
454	newleftblogs.blogs...	0		LabeledManually	45895.515282
387	madkane.com/not...	0		LeftyDirectory,Blo...	45021.616145
1479	wizbangblog.com	1		BlogPulse	40602.72767
1101	lashawnbarber.com	1		Blogarama,eTalkin...	36135.552523
1041	hughhewitt.com	1		Blogarama,Labele...	34249.665515
729	washingtonmonthl...	0		BlogPulse,LeftyDir...	32659.926605
363	liberaloasis.com	0		LeftyDirectory,Lab...	31778.007129
1153	michellemalkin.co...	1		BlogPulse	31411.488352
1000	newkaffeeanal.tune	1		LabeledManually	31398.343352

Betweenness Centrality đo vai trò trung gian của một nút trong mạng, tức là khả năng kết nối các nhóm hoặc cụm khác nhau. Nút có Betweenness cao đóng vai trò "cầu nối" trong việc truyền thông tin giữa các phần của mạng.

- [blogsforbush.com](#): Không chỉ có nhiều kết nối (Degree cao), mà còn đóng vai trò là "cầu nối" quan trọng trong mạng, giúp kết nối các cụm hoặc nhóm blog khác nhau.
- [atrios.blogspot.com](#): Là trung gian cho nhiều con đường thông tin trong mạng, giúp liên kết các phần tử không kết nối trực tiếp.
- [instapundit.com](#): Vừa có nhiều kết nối, vừa đóng vai trò quan trọng trong việc duy trì dòng chảy thông tin giữa các nhóm.

3. Closeness Centrality

Id	Label	Interval	value	source	Closeness Centrality ▾
820	batr.net		1	Blogarama	1.0
76	blog.dennisfox.net		0	eTalkingHead	1.0
245	greaterdemocracy....		0	CampaignLine	1.0
1330	scrappleface.com		1	LabeledManually	1.0
1293	right-thinking.com		1	eTalkingHead	1.0
898	cmillertime.blogspot...		1	eTalkingHead	1.0
1159	moorewatch.com		1	LabeledManually	1.0
588	scripting.com		0	BlogPulse	1.0
128	charleypatton.blog...		0	BlogCatalog,eTalki...	1.0
603	shrillblog.blogspot...		0	LabeledManually	1.0
745	wsrblood.typepad....		0	Blogarama,LeftyDi...	1.0
1468	watchandwait.blog...		1	BlogCatalog	1.0
141	connptions.net		0	Blogarama	1.0
1442	tuckereskew.typep...		1	CampaignLine	1.0
643	tallone.org		0	Blogarama	1.0
797	amradio.blogspot....		1	Blogarama	1.0

Closeness Centrality đo khả năng tiếp cận của một nút đến các nút khác trong mạng, thông qua các con đường ngắn nhất. Giá trị cao cho thấy nút đó có vị trí thuận lợi trong việc tiếp cận hoặc phát tán thông tin.

Các nút như batr.net, greaterdemocracy.org hay scrappleface.com đều có thể dễ dàng tiếp cận hoặc ảnh hưởng đến các phần tử khác trong mạng.

Phần 2: Phát hiện cộng đồng Thực hiện phân cụm mạng lưới sử dụng 3 thuật toán

sau:

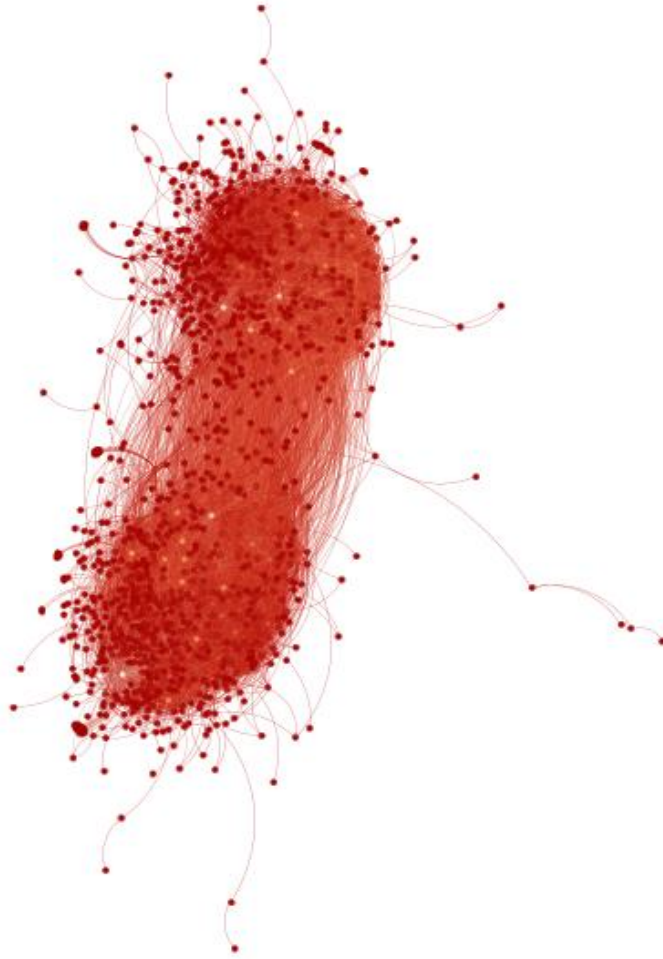
1. Thuật toán Louvain

Results:

Modularity: 0.427

Modularity with resolution: 0.427

Number of Communities: 277



2. Thuật toán Girvan-newman

3. Thuật toán LPA (gephi.org/plugins/#/plugin/label-propagation-clustering)