label propagation

Matthew McAvoy

September 9, 2019

Contents

TODO

- constant p and k=(4,6,8)
- background noise k=(4,6,8), nq=(100,200,400)

python mirai.py –communities 4 –background$_{\text{num}}$ 400 –background$_{\text{prob}}$ 0.01 0.51 python mirai.py –communities 6 –background$_{\text{num}}$ 100 –background$_{\text{prob}}$ 0.01 0.51 python mirai.py –communities 6 –background$_{\text{num}}$ 200 –background$_{\text{prob}}$ 0.01 0.51 python mirai.py –communities 6 –background$_{\text{num}}$ 400 –background$_{\text{prob}}$ 0.01 0.51 python mirai.py –communities 8 –background$_{\text{num}}$ 100 –background$_{\text{prob}}$ 0.01 0.51 python mirai.py –communities 8 –background$_{\text{num}}$ 200 –background$_{\text{prob}}$ 0.01 0.51 python mirai.py –communities 8 –background$_{\text{num}}$ 400 –background$_{\text{prob}}$ 0.01 0.51

- background noise flipping communities graphic python mirai.py –communities 2 –$\mathrm{background}_{\mathrm{num}}$ 100 –draw python mirai.py –communities 2 –$\mathrm{background}_{\mathrm{num}}$ 200 –draw python mirai.py –communities 2 –$\mathrm{background}_{\mathrm{num}}$ 400 –draw

1 Constant p values

How does label propagation work when p and p' are kept constant calculated from k=2 for larger k values. Expect once the number of connections from out of community is greater than inter community then it will fail

$p = 0.543, p' = 0.015, c = 0.00044, n = 100, samples = 10$

Expected number of neighbors: n * p1 or n * p2 * (k - 1) community neighbors: 100 * .543 = 54.3 outside neighbors: 100 * 0.015 * (51 - 1) = 61

We see with p' 0.1 and p held constant at n=100 per each community and k = 6 communities the algorithm sharply starts to fail. There are $E[n_p]$ = 54 and $E[n_p']$ = 60; but not when $E[n_p']$ = 50. This suggests as long as there are more edges within a community than between communities, the algorithm may still succeed.

index	n	p1	p2	nq	q	k	$\mathrm{path}_{\mathrm{length}}$	converged
0	100	0.543122	0.015364	0	0	1	1.455657	1.0
1	100	0.543122	0.015364	0	0	2	1.829131	1.0
2	100	0.543122	0.015364	0	0	3	1.930076	1.0
3	100	0.543122	0.015364	0	0	4	1.984492	1.0
4	100	0.543122	0.015364	0	0	5	2.024196	1.0
5	100	0.543122	0.015364	0	0	6	2.037989	1.0
6	100	0.543122	0.015364	0	0	7	2.051492	1.0
7	100	0.543122	0.015364	0	0	8	2.060753	1.0
8	100	0.543122	0.015364	0	0	9	2.068634	1.0
9	100	0.543122	0.015364	0	0	10	2.072378	1.0

index	n	p1	p2	nq	q	k	$\mathrm{path}_{\mathrm{length}}$	converged
0	100	0.543122	0.1	0	0	1	1.461313	1.0
1	100	0.543122	0.1	0	0	2	1.678859	1.0
2	100	0.543122	0.1	0	0	3	1.753817	1.0
3	100	0.543122	0.1	0	0	4	1.789917	1.0
4	100	0.543122	0.1	0	0	5	1.812291	0.9
5	100	0.543122	0.1	0	0	6	1.826776	0.8
6	100	0.543122	0.1	0	0	7	1.836800	0.0
7	100	0.543122	0.1	0	0	8	1.845112	0.1
8	100	0.543122	0.1	0	0	9	1.851191	0.0
9	100	0.543122	0.1	0	0	10	1.856121	0.0

2 p' value from two to one community

What is the p' value in which two distinct clusters becomes one cluster?

$p = 0.543, k = 2, c1 = 0.00044, c2 = n = 100, samples = 10$

index	p	p'	number of communities	average path length	label propagation
0	0.543122	0.00	2	0.000000	1.0
1	0.543122	0.01	2	1.902497	1.0
2	0.543122	0.02	2	1.779050	1.0
3	0.543122	0.03	2	1.731678	1.0
4	0.543122	0.04	2	1.718714	1.0
5	0.543122	0.05	2	1.705829	1.0
6	0.543122	0.06	2	1.699312	1.0
7	0.543122	0.07	2	1.694307	1.0
8	0.543122	0.08	2	1.690161	1.0
9	0.543122	0.09	2	1.685060	1.0
10	0.543122	0.10	2	1.677528	1.0
11	0.543122	0.11	2	1.675392	1.0
12	0.543122	0.12	2	1.669497	1.0
13	0.543122	0.13	2	1.665663	0.9
14	0.543122	0.14	2	1.660236	1.0
15	0.543122	0.15	2	1.652548	1.0
16	0.543122	0.16	2	1.649070	1.0
17	0.543122	0.17	2	1.644985	1.0
18	0.543122	0.18	2	1.639307	1.0
19	0.543122	0.19	2	1.634573	0.9
20	0.543122	0.20	2	1.629131	0.6
21	0.543122	0.21	2	1.625372	0.6
22	0.543122	0.22	2	1.618211	0.4
23	0.543122	0.23	2	1.614935	0.2
24	0.543122	0.24	2	1.607985	0.3
25	0.543122	0.25	2	1.604482	0.0
26	0.543122	0.26	2	1.599533	0.2
27	0.543122	0.27	2	1.594352	0.1
28	0.543122	0.28	2	1.589915	0.1
29	0.543122	0.29	2	1.582905	0.1
30	0.543122	0.30	2	1.579739	0.1
31	0.543122	0.31	2	1.575176	0.0
32	0.543122	0.32	2	1.568593	0.0
33	0.543122	0.33	2	1.564342	0.0
34	0.543122	0.34	2	1.559814	0.0
35	0.543122	0.35	2	1.552513	0.0
36	0.543122	0.36	2	1.551452	0.0
37	0.543122	0.37	2	1.543281	0.0
38	0.543122	0.38	2	1.540618	0.0
39	0.543122	0.39	2	1.534538	0.0
40	0.543122	0.40	2	1.528744	0.0
41	0.543122	0.41	2	1.524613	0.0
42	0.543122	0.42	2	1.519055	0.0
43	0.543122	0.43	2	1.515693	0.0
44	0.543122	0.44	2	1.509246	0.0
45	0.543122	0.45	2	1.504925	0.0
46	0.543122	0.46	2	1.501392	0.0
47	0.543122	0.47	2	1.492553	0.0

n	p1	p2	nq	q	k	$path_{length}$	converged
100	0.5431215481815791	0.0	0	0.0	4		1.0
100	0.5431215481815791	0.01	0	0.0	4	2.1034085213032583	1.0
100	0.5431215481815791	0.02	0	0.0	4	1.9288746867167919	1.0
100	0.5431215481815791	0.03	0	0.0	4	1.8665989974937343	1.0
100	0.5431215481815791	0.04	0	0.0	4	1.8421641604010026	1.0
100	0.5431215481815791	0.05	0	0.0	4	1.829696741854637	1.0
100	0.5431215481815791	0.06	0	0.0	4	1.8204448621553884	1.0
100	0.5431215481815791	0.07	0	0.0	4	1.8128834586466165	1.0
100	0.5431215481815791	0.08	0	0.0	4	1.8048095238095239	1.0
100	0.5431215481815791	0.09	0	0.0	4	1.7971679197994987	1.0
100	0.5431215481815791	0.1	0	0.0	4	1.7898408521303257	1.0
100	0.5431215481815791	0.11	0	0.0	4	1.7833483709273181	0.9
100	0.5431215481815791	0.12	0	0.0	4	1.7758383458646616	0.9
100	0.5431215481815791	0.13	0	0.0	4	1.7675551378446115	0.8
100	0.5431215481815791	0.14	0	0.0	4	1.759300751879699	0.6
100	0.5431215481815791	0.15	0	0.0	4	1.7521804511278194	0.6
100	0.5431215481815791	0.16	0	0.0	4	1.7455513784461154	0.3
100	0.5431215481815791	0.17	0	0.0	4	1.7376954887218043	0.0
100	0.5431215481815791	0.18	0	0.0	4	1.7303395989974937	0.0
100	0.5431215481815791	0.19	0	0.0	4	1.7231942355889724	0.0
100	0.5431215481815791	0.2	0	0.0	4	1.715687969924812	0.0
100	0.5431215481815791	0.21	0	0.0	4	1.7070401002506266	0.0
100	0.5431215481815791	0.22	0	0.0	4	1.6999348370927319	0.0
100	0.5431215481815791	0.23	0	0.0	4	1.6928659147869674	0.0
100	0.5431215481815791	0.24	0	0.0	4	1.6838320802005016	0.0
100	0.5431215481815791	0.25	0	0.0	4	1.6769987468671679	0.0
100	0.5431215481815791	0.26	0	0.0	4	1.6699335839598999	0.0
100	0.5431215481815791	0.27	0	0.0	4	1.662562656641604	0.0
100	0.5431215481815791	0.28	0	0.0	4	1.6544749373433583	0.0
100	0.5431215481815791	0.29	0	0.0	4	1.6462230576441104	0.0
100	0.5431215481815791	0.3	0	0.0	4	1.6387894736842104	0.0
100	0.5431215481815791	0.31	0	0.0	4	1.6313934837092734	0.0
100	0.5431215481815791	0.32	0	0.0	4	1.6241528822055138	0.0
100	0.5431215481815791	0.33	0	0.0	4	1.6167919799498747	0.0
100	0.5431215481815791	0.34	0	0.0	4	1.6094135338345865	0.0
100	0.5431215481815791	0.35000000000000003	0	0.0	4	1.6024799498746867	0.0
100	0.5431215481815791	0.36	0	0.0	4	1.5944010025062656	0.0
100	0.5431215481815791	0.37	0	0.0	4	1.586923558897243	0.0
100	0.5431215481815791	0.38	0	0.0	4	1.5785137844611528	0.0
100	0.5431215481815791	0.39	0	0.0	4	1.572588972431078	0.0
100	0.5431215481815791	0.4	0	0.0	4	1.563987468671679	0.0
100	0.5431215481815791	0.41000000000000003	0	0.0	4	1.5570238095238094	0.0
100	0.5431215481815791	0.42	0	0.0	4	1.5494724310776942	0.0
100	0.5431215481815791	0.43	0	0.0	4	1.5421002506265666	0.0
100	0.5431215481815791	0.44	0	0.0	4	1.534077694235589	0.0
100	0.5431215481815791	0.45	0	0.0	4	1.5265012531328321	0.0
100	0.5431215481815791	0.46	0	0.0	4	1.518645363408521	0.0
100	0.5431215481815791	0.47000000000000003	0	0.0	4	1.5114185463659147	0.0

n	p1	p2	nq	q	k	$\text{path}_{\text{length}}$	converged
100	0.5431215481815791	0.0	0	0.0	6		1.0
100	0.5431215481815791	0.01	0	0.0	6	2.1720623260990544	1.0
100	0.5431215481815791	0.02	0	0.0	6	1.973797996661102	1.0
100	0.5431215481815791	0.03	0	0.0	6	1.9075870895937672	1.0
100	0.5431215481815791	0.04	0	0.0	6	1.8823411240957149	1.0
100	0.5431215481815791	0.05	0	0.0	6	1.8693088480801336	1.0
100	0.5431215481815791	0.06	0	0.0	6	1.8605358931552587	1.0
100	0.5431215481815791	0.07	0	0.0	6	1.851744017807457	1.0
100	0.5431215481815791	0.08	0	0.0	6	1.8432921535893154	1.0
100	0.5431215481815791	0.09	0	0.0	6	1.8350072342793542	0.9
100	0.5431215481815791	0.1	0	0.0	6	1.8266683361157487	0.8
100	0.5431215481815791	0.11	0	0.0	6	1.8183405676126878	0.6
100	0.5431215481815791	0.12	0	0.0	6	1.8099321090706735	0.1
100	0.5431215481815791	0.13	0	0.0	6	1.8020617696160268	0.0
100	0.5431215481815791	0.14	0	0.0	6	1.7926232609905397	0.0
100	0.5431215481815791	0.15	0	0.0	6	1.7846071229827491	0.0
100	0.5431215481815791	0.16	0	0.0	6	1.7758942682248193	0.0
100	0.5431215481815791	0.17	0	0.0	6	1.7683388981636061	0.0
100	0.5431215481815791	0.18	0	0.0	6	1.7604396215915414	0.0
100	0.5431215481815791	0.19	0	0.0	6	1.7519632721202005	0.0
100	0.5431215481815791	0.2	0	0.0	6	1.7428842515303284	0.0
100	0.5431215481815791	0.21	0	0.0	6	1.7351463550361714	0.0
100	0.5431215481815791	0.22	0	0.0	6	1.7266928213689483	0.0
100	0.5431215481815791	0.23	0	0.0	6	1.7180934891485808	0.0
100	0.5431215481815791	0.24	0	0.0	6	1.710339454646633	0.0
100	0.5431215481815791	0.25	0	0.0	6	1.7016622148024485	0.0
100	0.5431215481815791	0.26	0	0.0	6	1.6925581524763493	0.0
100	0.5431215481815791	0.27	0	0.0	6	1.684870895937674	0.0
100	0.5431215481815791	0.28	0	0.0	6	1.6766382860322762	0.0
100	0.5431215481815791	0.29	0	0.0	6	1.6677417918753477	0.0
100	0.5431215481815791	0.3	0	0.0	6	1.6591969949916527	0.0

n	p1	p2	nq	q	k	$path_{length}$	converged
100	0.5431215481815791	0.0	0	0.0	8		1.0
100	0.5431215481815791	0.01	0	0.0	8	2.200523779724656	1.0
100	0.5431215481815791	0.02	0	0.0	8	1.9924721526908633	1.0
100	0.5431215481815791	0.03	0	0.0	8	1.9256673967459323	1.0
100	0.5431215481815791	0.04	0	0.0	8	1.902060700876095	1.0
100	0.5431215481815791	0.05	0	0.0	8	1.8897127659574469	1.0
100	0.5431215481815791	0.06	0	0.0	8	1.8801764705882351	1.0
100	0.5431215481815791	0.07	0	0.0	8	1.871589799749687	0.9
100	0.5431215481815791	0.08	0	0.0	8	1.8626968085106383	0.9
100	0.5431215481815791	0.09	0	0.0	8	1.8537390488110141	0.4
100	0.5431215481815791	0.1	0	0.0	8	1.8450575719649565	0.1
100	0.5431215481815791	0.11	0	0.0	8	1.8362753441802255	0.0
100	0.5431215481815791	0.12	0	0.0	8	1.8280685231539422	0.0
100	0.5431215481815791	0.13	0	0.0	8	1.8188204005006259	0.0
100	0.5431215481815791	0.14	0	0.0	8	1.8100797872340426	0.0
100	0.5431215481815791	0.15	0	0.0	8	1.801576032540676	0.0
100	0.5431215481815791	0.16	0	0.0	8	1.7923895494367958	0.0
100	0.5431215481815791	0.17	0	0.0	8	1.7837540675844807	0.0
100	0.5431215481815791	0.18	0	0.0	8	1.7751645807259073	0.0
100	0.5431215481815791	0.19	0	0.0	8	1.7660419274092614	0.0
100	0.5431215481815791	0.2	0	0.0	8	1.7570503754693365	0.0
100	0.5431215481815791	0.21	0	0.0	8	1.7485059449311637	0.0
100	0.5431215481815791	0.22	0	0.0	8	1.7400804130162704	0.0
100	0.5431215481815791	0.23	0	0.0	8	1.7312894242803503	0.0
100	0.5431215481815791	0.24	0	0.0	8	1.7220309762202757	0.0
100	0.5431215481815791	0.25	0	0.0	8	1.7136933667083851	0.0
100	0.5431215481815791	0.26	0	0.0	8	1.7051223404255318	0.0
100	0.5431215481815791	0.27	0	0.0	8	1.6961088861076346	0.0
100	0.5431215481815791	0.28	0	0.0	8	1.6875885481852315	0.0
100	0.5431215481815791	0.29	0	0.0	8	1.6789033166458072	0.0
100	0.5431215481815791	0.3	0	0.0	8	1.6696423654568207	0.0

3 p' graphs

$p = 0.543, p' = p/2 = .271, k = 2$

directory $p2_{k2success}$

A high p' value acquired the correct label 1/20 times. Noticed on the third step one of the communities had nearly completed acquiring its predicted label and the other community had a backbone that had also acquired the correct label. After 8 steps the second community finished updating it's correct label while the first community remained mostly unchanged in this time.

From theorem II they proved $X'' - Y'' > 0$. Border case is when $X' - Y' = 0$, $X'' - Y'' = 0$, and perhaps as well $X''' - Y''' = 0$,

$X = n_i \beta p_i \qquad Y = n_i p_i (1 - \frac{3p_i}{4}) \beta$

$X' = X - \sqrt{3cX \log n} \qquad Y' = Y + \sqrt{3cY \log n}$

$X' - Y' = 0 => X' = Y'$

$X - \sqrt{3cX \log n} = Y + \sqrt{3cY \log n}$

$n_i \beta p_i - \sqrt{3cn_i \beta p_i \log n} = n_i p_i (1 - \frac{3p_i}{4}) \beta + \sqrt{3cn_i p_i (1 - \frac{3p_i}{4}) \beta \log n}$

Nearly uncomputable. Proof used X > Y, perhaps look at when Y > X? Also consider *beta* is a larger value. Could it be on each step we require X > Y?

4 background noise

Is the label propagation algorithm still able to work when there are nodes not part of a community? Iterate over different q values (background connectivity) and $\mathrm{n_q}$ is the number of nodes in the background

$p = 0.543, p' = 0.015, k = 2, c1 = 0.00044, c2 = n = 100, trials = 10, steps = 50$

Can see an example where \$q = p2, $\mathrm{n_q}$ \$ in directory results/noise/

Hypothesis, for $nq = any$ and $q <= p'$ it should work Hypothesis, for $nq < c$ and $q < p$ it might

Find the background can be quite larger than the p' value and still propagate. However did find that after a pretty long time (typically converges in 3-4 steps) after 20 steps the labels in the background will converge then start to influence the communities

Perhaps we can use a log n stopping time to identify the communities and stop before noise in the background overwhelms the communities

Additionally see the algorithm may still converge, but requires less connectivity to the noise as the number of nodes in the background increases. Suggests there is a constant around E[$\mathrm{n_q}$] = 30 that is a threshold.

Expected number of edges for a node in a community connected to other nodes in a community is greater than the number of edges to nodes in background nq 100 k 2 converge fails at q 0.30 E[$\mathrm{n_q}$] = 30 < E[$\mathrm{n_p}$] = 54 nq 200 k 2 converge fails at q 0.15 E[$\mathrm{n_q}$] = 30 < E[$\mathrm{n_p}$] = 54 nq 300 k 2 converge fails at q 0.10 E[$\mathrm{n_q}$] = 30 < E[$\mathrm{n_p}$] = 54 nq 400 k 2 converge fails at q 0.06 E[$\mathrm{n_q}$] = 26 < E[$\mathrm{n_p}$] = 54 nq 100 k 4 converge fails at q 0.36 E[$\mathrm{n_q}$] = 36 < E[$\mathrm{n_p}$] =

54 nq 200 k 4 converge fails at q 0.17 $\mathrm{E}[\mathrm{n_q}] = 34 < \mathrm{E}[\mathrm{n_p}] = 54$ nq 400 k 4 converge fails at q 0.08 $\mathrm{E}[\mathrm{n_q}] = 32 < \mathrm{E}[\mathrm{n_p}] = 54$

n	p1	p2	nq	q	k	$path_{length}$	converged
100	0.5431215481815791	0.1	100	0.0	2	0.0	1.0
100	0.5431215481815791	0.1	100	0.01	2	0.0	1.0
100	0.5431215481815791	0.1	100	0.02	2	1.8622385730211817	1.0
100	0.5431215481815791	0.1	100	0.03	2	1.977201783723523	1.0
100	0.5431215481815791	0.1	100	0.04	2	1.930015607580825	1.0
100	0.5431215481815791	0.1	100	0.05	2	1.894488294314381	1.0
100	0.5431215481815791	0.1	100	0.06	2	1.8687402452619846	1.0
100	0.5431215481815791	0.1	100	0.07	2	1.8449052396878483	1.0
100	0.5431215481815791	0.1	100	0.08	2	1.8292151616499441	1.0
100	0.5431215481815791	0.1	100	0.09	2	1.8186332218506132	1.0
100	0.5431215481815791	0.1	100	0.1	2	1.8081605351170567	1.0
100	0.5431215481815791	0.1	100	0.11	2	1.7993266443701224	1.0
100	0.5431215481815791	0.1	100	0.12	2	1.791386845039019	1.0
100	0.5431215481815791	0.1	100	0.13	2	1.7867870680044593	1.0
100	0.5431215481815791	0.1	100	0.14	2	1.7797837235228542	1.0
100	0.5431215481815791	0.1	100	0.15	2	1.7758818283166111	1.0
100	0.5431215481815791	0.1	100	0.16	2	1.7678305462653288	1.0
100	0.5431215481815791	0.1	100	0.17	2	1.7630880713489412	1.0
100	0.5431215481815791	0.1	100	0.18	2	1.7568071348940915	0.9
100	0.5431215481815791	0.1	100	0.19	2	1.7513311036789296	1.0
100	0.5431215481815791	0.1	100	0.2	2	1.7455741360089185	1.0
100	0.5431215481815791	0.1	100	0.21	2	1.7406510590858417	1.0
100	0.5431215481815791	0.1	100	0.22	2	1.7354604236343367	1.0
100	0.5431215481815791	0.1	100	0.23	2	1.7302987736900781	1.0
100	0.5431215481815791	0.1	100	0.24	2	1.723346711259755	0.7
100	0.5431215481815791	0.1	100	0.25	2	1.7184080267558528	1.0
100	0.5431215481815791	0.1	100	0.26	2	1.7147357859531773	0.5
100	0.5431215481815791	0.1	100	0.27	2	1.7068338907469343	0.4
100	0.5431215481815791	0.1	100	0.28	2	1.7024860646599778	0.5
100	0.5431215481815791	0.1	100	0.29	2	1.6975785953177258	0.1
100	0.5431215481815791	0.1	100	0.3	2	1.6920178372352286	0.1
100	0.5431215481815791	0.1	100	0.31	2	1.684882943143813	0.1
100	0.5431215481815791	0.1	100	0.32	2	1.6805039018952062	0.0
100	0.5431215481815791	0.1	100	0.33	2	1.6740022296544033	0.0
100	0.5431215481815791	0.1	100	0.34	2	1.6677591973244144	0.0
100	0.5431215481815791	0.1	100	0.35000000000000003	2	1.6633979933110368	0.0
100	0.5431215481815791	0.1	100	0.36	2	1.6573623188405797	0.0
100	0.5431215481815791	0.1	100	0.37	2	1.6516432552954292	0.0
100	0.5431215481815791	0.1	100	0.38	2	1.6475496098104796	0.0
100	0.5431215481815791	0.1	100	0.39	2	1.6415830546265329	0.0
100	0.5431215481815791	0.1	100	0.4	2	1.6359754738015606	0.0
100	0.5431215481815791	0.1	100	0.41000000000000003	2	1.6298862876254183	0.0
100	0.5431215481815791	0.1	100	0.42	2	1.6245641025641024	0.0
100	0.5431215481815791	0.1	100	0.43	2	1.618577480490524	0.0
100	0.5431215481815791	0.1	100	0.44	2	1.612191750278707	0.0
100	0.5431215481815791	0.1	100	0.45	2	1.6065841694537348	0.0
100	0.5431215481815791	0.1	100	0.46	2	1.6014358974358973	0.0
100	0.5431215481815791	0.1	100	0.47000000000000003	2	1.596608695652174	0.0

n	p1	p2	nq	q	k	pat
100	0.5431215481815791	0.015363594588497674	100	0.01	4	2.1095807615
100	0.5431215481815791	0.015363594588497674	100	0.02	4	2.108452905
100	0.5431215481815791	0.015363594588497674	100	0.03	4	2.0467527054
100	0.5431215481815791	0.015363594588497674	100	0.04	4	2.0065306613
100	0.5431215481815791	0.015363594588497674	100	0.05	4	1.9751967935
100	0.5431215481815791	0.015363594588497674	100	0.060000000000000005	4	1.9570484969
100	0.5431215481815791	0.015363594588497674	100	0.06999999999999999	4	1.940113827
100	0.5431215481815791	0.015363594588497674	100	0.08	4	1.9257458917
100	0.5431215481815791	0.015363594588497674	100	0.09	4	1.9127607214
100	0.5431215481815791	0.015363594588497674	100	0.09999999999999999	4	1.90183246
100	0.5431215481815791	0.015363594588497674	100	0.11	4	1.8928521042
100	0.5431215481815791	0.015363594588497674	100	0.12	4	1.882685370
100	0.5431215481815791	0.015363594588497674	100	0.13	4	1.8754004008
100	0.5431215481815791	0.015363594588497674	100	0.14	4	1.867397194
100	0.5431215481815791	0.015363594588497674	100	0.15000000000000002	4	1.8612617234
100	0.5431215481815791	0.015363594588497674	100	0.16	4	1.8553963927
100	0.5431215481815791	0.015363594588497674	100	0.17	4	1.8500953907
100	0.5431215481815791	0.015363594588497674	100	0.18000000000000002	4	1.8446356713
100	0.5431215481815791	0.015363594588497674	100	0.19	4	1.8405442885
100	0.5431215481815791	0.015363594588497674	100	0.2	4	1.8358108216
100	0.5431215481815791	0.015363594588497674	100	0.21000000000000002	4	1.8316705410
100	0.5431215481815791	0.015363594588497674	100	0.22	4	1.8274228456
100	0.5431215481815791	0.015363594588497674	100	0.23	4	1.8235118236
100	0.5431215481815791	0.015363594588497674	100	0.24000000000000002	4	1.8209402805
100	0.5431215481815791	0.015363594588497674	100	0.25	4	1.81614749
100	0.5431215481815791	0.015363594588497674	100	0.26	4	1.8123687374
100	0.5431215481815791	0.015363594588497674	100	0.27	4	1.8092977955
100	0.5431215481815791	0.015363594588497674	100	0.28	4	1.8056745490
100	0.5431215481815791	0.015363594588497674	100	0.29000000000000004	4	1.801774749
100	0.5431215481815791	0.015363594588497674	100	0.3	4	1.7977651302
100	0.5431215481815791	0.015363594588497674	100	0.31	4	1.7946509018
100	0.5431215481815791	0.015363594588497674	100	0.32	4	1.7913498997
100	0.5431215481815791	0.015363594588497674	100	0.33	4	1.7875711422
100	0.5431215481815791	0.015363594588497674	100	0.34	4	1.7840617234
100	0.5431215481815791	0.015363594588497674	100	0.35000000000000003	4	1.7805851703
100	0.5431215481815791	0.015363594588497674	100	0.36000000000000004	4	1.7774669338
100	0.5431215481815791	0.015363594588497674	100	0.37	4	1.7730725450
100	0.5431215481815791	0.015363594588497674	100	0.38	4	1.769163126
100	0.5431215481815791	0.015363594588497674	100	0.39	4	1.7659462925
100	0.5431215481815791	0.015363594588497674	100	0.4	4	1.7621130260
100	0.5431215481815791	0.015363594588497674	100	0.41000000000000003	4	1.7583895791
100	0.5431215481815791	0.015363594588497674	100	0.42000000000000004	4	1.7554420841
100	0.5431215481815791	0.015363594588497674	100	0.43	4	1.7511070140
100	0.5431215481815791	0.015363594588497674	100	0.44	4	1.747834869
100	0.5431215481815791	0.015363594588497674	100	0.45	4	1.7438557114
100	0.5431215481815791	0.015363594588497674	100	0.46	4	1.7402164328
100	0.5431215481815791	0.015363594588497674	100	0.47000000000000003	4	1.7367118236
100	0.5431215481815791	0.015363594588497674	100	0.48000000000000004	4	1.733244088

n	p1	p2	nq	q	k	pat
100	0.5431215481815791	0.015363594588497674	400	0.01	4	2.5105522528
100	0.5431215481815791	0.015363594588497674	400	0.02	4	2.2849292866
100	0.5431215481815791	0.015363594588497674	400	0.03	4	2.14046464
100	0.5431215481815791	0.015363594588497674	400	0.04	4	2.042555381
100	0.5431215481815791	0.015363594588497674	400	0.05	4	1.9789198998
100	0.5431215481815791	0.015363594588497674	400	0.060000000000000005	4	1.942862640
100	0.5431215481815791	0.015363594588497674	400	0.06999999999999999	4	1.9211311013
100	0.5431215481815791	0.015363594588497674	400	0.08	4	1.9077124530
100	0.5431215481815791	0.015363594588497674	400	0.09	4	1.8976091989
100	0.5431215481815791	0.015363594588497674	400	0.09999999999999999	4	1.889011576
100	0.5431215481815791	0.015363594588497674	400	0.11	4	1.8811971214
100	0.5431215481815791	0.015363594588497674	400	0.12	4	1.873470588
100	0.5431215481815791	0.015363594588497674	400	0.13	4	1.865758760
100	0.5431215481815791	0.015363594588497674	400	0.14	4	1.858307259
100	0.5431215481815791	0.015363594588497674	400	0.15000000000000002	4	1.8509812265
100	0.5431215481815791	0.015363594588497674	400	0.16	4	1.8436270337
100	0.5431215481815791	0.015363594588497674	400	0.17	4	1.835861076
100	0.5431215481815791	0.015363594588497674	400	0.18000000000000002	4	1.8282478097
100	0.5431215481815791	0.015363594588497674	400	0.19	4	1.821177096
100	0.5431215481815791	0.015363594588497674	400	0.2	4	1.8133301001
100	0.5431215481815791	0.015363594588497674	400	0.21000000000000002	4	1.8060982478
100	0.5431215481815791	0.015363594588497674	400	0.22	4	1.7982543804
100	0.5431215481815791	0.015363594588497674	400	0.23	4	1.7907340425
100	0.5431215481815791	0.015363594588497674	400	0.24000000000000002	4	1.7839105131
100	0.5431215481815791	0.015363594588497674	400	0.25	4	1.7757046307
100	0.5431215481815791	0.015363594588497674	400	0.26	4	1.7684724655
100	0.5431215481815791	0.015363594588497674	400	0.27	4	1.7609834167
100	0.5431215481815791	0.015363594588497674	400	0.28	4	1.7536833541
100	0.5431215481815791	0.015363594588497674	400	0.29000000000000004	4	1.7462518773
100	0.5431215481815791	0.015363594588497674	400	0.3	4	1.7384755944
100	0.5431215481815791	0.015363594588497674	400	0.31	4	1.7311170212
100	0.5431215481815791	0.015363594588497674	400	0.32	4	1.7231799123
100	0.5431215481815791	0.015363594588497674	400	0.33	4	1.7156761576
100	0.5431215481815791	0.015363594588497674	400	0.34	4	1.7084195869
100	0.5431215481815791	0.015363594588497674	400	0.35000000000000003	4	1.7008081977
100	0.5431215481815791	0.015363594588497674	400	0.36000000000000004	4	1.6933911138
100	0.5431215481815791	0.015363594588497674	400	0.37	4	1.6860563204
100	0.5431215481815791	0.015363594588497674	400	0.38	4	1.6787490613
100	0.5431215481815791	0.015363594588497674	400	0.39	4	1.6709705882
100	0.5431215481815791	0.015363594588497674	400	0.4	4	1.6636035669
100	0.5431215481815791	0.015363594588497674	400	0.41000000000000003	4	1.6561617647
100	0.5431215481815791	0.015363594588497674	400	0.42000000000000004	4	1.648203692
100	0.5431215481815791	0.015363594588497674	400	0.43	4	1.6412180851
100	0.5431215481815791	0.015363594588497674	400	0.44	4	1.6331210888
100	0.5431215481815791	0.015363594588497674	400	0.45	4	1.6263354192
100	0.5431215481815791	0.015363594588497674	400	0.46	4	1.6185306633
100	0.5431215481815791	0.015363594588497674	400	0.47000000000000003	4	1.610886420
100	0.5431215481815791	0.015363594588497674	400	0.48000000000000004	4	1.603078535

5 variable p values

Is it possible for label propagation algorithm to work in a network where there p values are different, resulting in a heterogeneous distribution of degrees with a constant p'?

$\mathrm{p_i}$ is generated from a uniform distribution between the indicated min p_i and 1.

Hypothesis, if min $p_i >$ theorem $2p$ for all $i \in V_1$ then it should Hypothesis, if min $p_i > p'$ for all $i \in V_1$ then it might Hypothesis, if min $p_i <= p'$ for all $i \in V_1$ then it should not

$ k = (2,4,6), nq = 0, p' = 0.01536, trials = 10, steps = 50 $

Observation: Once the minimum intra-community probability is larger than background probability, the algorithm is able to converge. This suggests as long as a community has a lower bound probability larger than background, the algorithm may be able to find the community. Its even possible for a range (0,1) intra prob for communities in a network to converge.

Code: python mirai.py –$\mathrm{intra_{prob}}$ 0.0 0.51 –$\mathrm{background_{num}}$ 100 –randomized

n	p1	p2	nq	q	k	$path_{length}$	converg
100	0.0	0.015363594588497674	0	0.015363594588497674	2	2.8391959798994977	
100	0.01	0.015363594588497674	0	0.015363594588497674	2	2.9834572864321607	
100	0.02	0.015363594588497674	0	0.015363594588497674	2	2.8366633165829147	
100	0.03	0.015363594588497674	0	0.015363594588497674	2	3.011874371859297	
100	0.04	0.015363594588497674	0	0.015363594588497674	2	2.8186331658291457	
100	0.05	0.015363594588497674	0	0.015363594588497674	2	2.977613065326633	
100	0.06	0.015363594588497674	0	0.015363594588497674	2	2.862366834170854	
100	0.07	0.015363594588497674	0	0.015363594588497674	2	2.726120603015075	
100	0.08	0.015363594588497674	0	0.015363594588497674	2	2.634984924623116	
100	0.09	0.015363594588497674	0	0.015363594588497674	2	2.556708542713568	
100	0.1	0.015363594588497674	0	0.015363594588497674	2	2.504236180904523	
100	0.11	0.015363594588497674	0	0.015363594588497674	2	2.430507537688442	
100	0.12	0.015363594588497674	0	0.015363594588497674	2	2.388462311557789	
100	0.13	0.015363594588497674	0	0.015363594588497674	2	2.343874371859296	
100	0.14	0.015363594588497674	0	0.015363594588497674	2	2.3061206030150756	
100	0.15	0.015363594588497674	0	0.015363594588497674	2	2.2861658291457285	
100	0.16	0.015363594588497674	0	0.015363594588497674	2	2.259819095477387	
100	0.17	0.015363594588497674	0	0.015363594588497674	2	2.2228341708542714	
100	0.18	0.015363594588497674	0	0.015363594588497674	2	2.2013065326633163	
100	0.19	0.015363594588497674	0	0.015363594588497674	2	2.1886231155778892	
100	0.2	0.015363594588497674	0	0.015363594588497674	2	2.174798994974874	
100	0.21	0.015363594588497674	0	0.015363594588497674	2	2.1468391959799	
100	0.22	0.015363594588497674	0	0.015363594588497674	2	2.1312211055276387	
100	0.23	0.015363594588497674	0	0.015363594588497674	2	2.1176532663316587	
100	0.24	0.015363594588497674	0	0.015363594588497674	2	2.1059296482412058	
100	0.25	0.015363594588497674	0	0.015363594588497674	2	2.0990402010050255	
100	0.26	0.015363594588497674	0	0.015363594588497674	2	2.0810603015075375	
100	0.27	0.015363594588497674	0	0.015363594588497674	2	2.0796532663316585	
100	0.28	0.015363594588497674	0	0.015363594588497674	2	2.060115577889447	
100	0.29	0.015363594588497674	0	0.015363594588497674	2	2.0473316582914576	
100	0.3	0.015363594588497674	0	0.015363594588497674	2	2.0415929648241207	
100	0.31	0.015363594588497674	0	0.015363594588497674	2	2.032743718592965	
100	0.32	0.015363594588497674	0	0.015363594588497674	2	2.020798994974874	
100	0.33	0.015363594588497674	0	0.015363594588497674	2	1.9995577889447236	
100	0.34	0.015363594588497674	0	0.015363594588497674	2	1.9949597989949748	
100	0.35	0.015363594588497674	0	0.015363594588497674	2	1.990643216080402	
100	0.36	0.015363594588497674	0	0.015363594588497674	2	1.975175879396985	
100	0.37	0.015363594588497674	0	0.015363594588497674	2	1.9683768844221103	
100	0.38	0.015363594588497674	0	0.015363594588497674	2	1.9638291457286432	
100	0.39	0.015363594588497674	0	0.015363594588497674	2	1.9419547738693468	
100	0.4	0.015363594588497674	0	0.015363594588497674	2	1.9404623115577888	
100	0.41	0.015363594588497674	0	0.015363594588497674	2	1.9265829145728641	
100	0.42	0.015363594588497674	0	0.015363594588497674	2	1.9243819095477388	
100	0.43	0.015363594588497674	0	0.015363594588497674	2	1.9099497487437183	
100	0.44	0.015363594588497674	0	0.015363594588497674	2	1.8983316582914571	
100	0.45	0.015363594588497674	0	0.015363594588497674	2	1.893256281407035	
100	0.46	0.015363594588497674	0	0.015363594588497674	2	1.8776482412060305	
100	0.47	0.015363594588497674	0	0.015363594588497674	2	1.8757839195979902	

6 Proof

Hypothesis: Assume theorems below. Given a network that satisfies them, a label propagation algorithm will be able to with high probability identify communities where intra-probability connectivity satisifies theorem I and n / inter-probability satisfies theorem II. This extends an normal Erdos-Renyi graph by allowing a variable intra-probability vs a static value and a background noise nodes that are not part of any community.

6.1 Node degree

- In a random graph (Erdos-Renyi) the node degree of any two nodes is uncorrelated
- Hypothesis: The average node degree of a node in a cluster is higher than out

6.2 Review of Theorem 2

Let $G(\Pi, \pi, p')$ be a clustered Erdos-Renyi graph. Suppose that the probabilities $\{p_i\}$ and p_i and the node subset sizes $\{n_i\}$ and n satisfy the inequalities

$$(i) n_i p_i^2 > 8np'$$

$$(ii) n_i p_i^4 > 1800c \log(n)$$

for some constant c. Then, given input $G(\Pi, \pi, p')$, Max-LPA converges correctly to node patition Π in two rounds w.h.p. (Note that condition (ii) implies for each i, $p_i > \frac{log(n_i)}{n_i}$

6.3 Review of bounds

With high probability (w.h.p.) $(1 - \frac{1}{n^c} \quad c \geq 1$ will converge

Let $X_1, X_2, ..., X_m$ be i.i.d. indicator random variables. Let $X = \sum_{i=1}^{m} X_i$. Use Chernoff and Hoffding bounds.

$$Pr(X > (1+\epsilon)E[X]) \leq e^{-\frac{\epsilon^2 E[X]}{3}}$$

$$Pr(X < (1-\epsilon)E[X]) \leq e^{-\frac{\epsilon^2 E[X]}{2}}$$

$$Pr(|X - E[X]| > \sqrt{3cE[X]\log(n)}) \leq \frac{1}{n^c}$$

6.4 Expected intra vs inter

As the labels of nodes are uniformly distributed, the probability a node acquires a label is uniformly distributed among its available edges. On the first round. Thus need to show the degree intra > degree inter.

What about on the second round? The average number of edges for any node intra that acquired a label from inta > inter.

If the expected number of nodes in a community recieve a label within its community is greater than an out-of-community member then after (log(n))? steps, will converge to a community detected.

expected intra-community

$$X := n_i p_i$$

Where n_i is the number of nodes in the community and p_i is the average probability in the community i.

$$Y := (n - n_i)p'$$

Where n is the total number of nodes in the network and p' is the background noise.

Thus on the first round

$$X - Y = n_i p_i - (n - n_i)p' > 0$$

Thus after the first round more nodes intra will acquire an intra label vs inter label

We also know

$$d_k > d'_k$$

On the second round and any round after

	community	outer
community	n_i p_i	$(n - n_i)$ p'
outer	n_i p'	$(n - n_i)$ p'