Urban Computing Assignment 2 – Saket Vishwasrao

1) The page rank and hits score for CA graph.

Node	Hits	Node	Pagerank
534751	0.118981	225438	2.41E-06
498390	0.082091	287362	2.39E-06
534753	0.068811	403664	2.11E-06
534779	0.065341	241926	2.07E-06
534777	0.061817	748883	2.07E-06
498389	0.057105	264566	2.01E-06
534774	0.040126	1898195	2.00E-06
534768	0.036013	156509	1.98E-06
534767	0.03134	1014744	1.95E-06
498404	0.030694	749665	1.95E-06
498391	0.027777	673507	1.93E-06
498220	0.023622	1766166	1.93E-06
498393	0.021803	737950	1.92E-06
534775	0.019688	241299	1.90E-06
534746	0.018045	1682326	1.90E-06
534750	0.014742	1362200	1.89E-06
498222	0.012592	1876000	1.87E-06
534765	0.012383	450536	1.85E-06
535446	0.012049	470832	1.85E-06
534764	0.011474	1258079	1.85E-06
534754	0.011309	1001106	1.82E-06
498221	0.011307	827039	1.82E-06
534766	0.010914	1972	1.82E-06
498333	0.010835	1962395	1.81E-06
534772	0.009722	1744403	1.81E-06
498392	0.009044	806073	1.78E-06
534776	0.008861	1709314	1.78E-06
498388	0.008753	1043614	1.78E-06
534778	0.00835	1695198	1.77E-06
534752	0.008159	695108	1.77E-06
498409	0.008089	128640	1.77E-06
498219	0.007541	1048758	1.77E-06
498405	0.006643	1514779	1.76E-06
498305	0.006568	612334	1.76E-06
534748	0.005798	1463627	1.76E-06
534769	0.005713	119583	1.75E-06
534773	0.004913	1609248	1.75E-06
498223	0.004722	1569945	1.74E-06

498408	0.004489	1661235	1.74E-06
498387	0.004161	811261	1.73E-06
498334	0.003361	1466612	1.73E-06
498341	0.003327	1889987	1.73E-06
498394	0.00329	1024958	1.73E-06
534756	0.003061	749926	1.72E-06
498416	0.002953	1195782	1.72E-06
498306	0.002646	416129	1.70E-06
498340	0.002638	50074	1.70E-06
498303	0.002589	1663362	1.70E-06
498217	0.002446	1066860	1.70E-06
498373	0.00228	1791732	1.70E-06
534744	0.002269	723604	1.69E-06
534749	0.002208	1022434	1.69E-06
498301	0.002126	342989	1.69E-06
498406	0.002086	1533807	1.69E-06
498193	0.002074	1569267	1.69E-06
498302	0.002063	553540	1.69E-06
498397	0.001947	318731	1.69E-06
498194	0.001651	1102535	1.68E-06
498199	0.001545	1464685	1.68E-06
498175	0.001337	255255	1.68E-06
498374	0.00119	977378	1.68E-06
498395	0.001158	441990	1.68E-06
498339	0.001152	241646	1.68E-06
534757	0.001085	1275439	1.68E-06
498414	0.001073	686122	1.68E-06
498216	0.001029	1467146	1.67E-06
534770	0.000952	1777991	1.67E-06
498407	0.000916	708301	1.67E-06
498224	0.000879	1428948	1.67E-06
534747	0.000865	1346523	1.67E-06
498345	0.000857	1461926	1.67E-06
498304	0.000799	1840885	1.67E-06
498300	0.000752	309759	1.67E-06
535445	0.000717	1606368	1.66E-06
498375	0.00066	1258479	1.66E-06
498192	0.000651	1963085	1.66E-06
498185	0.000549	322136	1.66E-06
498418	0.000541	145318	1.66E-06
498297	0.00053	1432477	1.66E-06
498415	0.000479	483262	1.66E-06
498424	0.000432	114101	1.66E-06

498172	0.000415	888498	1.66E-06
498207	0.000413	827511	1.66E-06
498198	0.000376	451272	1.66E-06
534745	0.000366	1791842	1.66E-06
498425	0.000356	826026	1.65E-06
498211	0.000356	1422209	1.65E-06
498396	0.000347	1381826	1.65E-06
498213	0.000326	585374	1.65E-06
534755	0.000324	1496146	1.65E-06
534742	0.000288	1652993	1.65E-06
498336	0.000279	562818	1.65E-06
498335	0.000244	1503646	1.65E-06
498180	0.00024	1063379	1.65E-06
498323	0.000239	1949624	1.65E-06
498419	0.000232	1512294	1.65E-06
498166	0.00022	1607775	1.65E-06
498225	0.00021	741233	1.65E-06
498200	0.000206	435933	1.65E-06
498307	0.000206	521168	1.65E-06

2) For the PA graph:

Node	Hits	Node	Pagerank
754877	0.142568	1025812	4.28E-06
754996	0.132254	268678	3.87E-06
754873	0.112211	40165	3.77E-06
754999	0.102816	721623	3.69E-06
754948	0.073238	123515	3.69E-06
755000	0.052615	868283	3.55E-06
754912	0.050253	124056	3.53E-06
754876	0.050128	63289	3.50E-06
754997	0.035591	859326	3.41E-06
754869	0.030933	559800	3.38E-06
754949	0.030928	134245	3.37E-06
754995	0.023152	536360	3.31E-06
754950	0.011605	532292	3.26E-06
754951	0.011206	623175	3.26E-06
755004	0.010713	131585	3.23E-06
754872	0.010687	223521	3.21E-06
755002	0.010561	304158	3.19E-06

778545	0.00805	712341	3.19E-06
755001	0.007	438489	3.19E-06
754875	0.004887	102420	3.18E-06
754871	0.004506	136445	3.16E-06
754874	0.004376	242967	3.15E-06
755008	0.003603	513173	3.13E-06
778567	0.003307	411160	3.11E-06
755003	0.003188	267891	3.11E-06
754908	0.003139	767711	3.10E-06
374895	0.002778	62471	3.10E-06
752987	0.002444	167635	3.08E-06
374908	0.002393	466256	3.08E-06
374896	0.002215	769107	3.08E-06
764228	0.002161	508545	3.08E-06
375728	0.0019	507048	3.07E-06
754862	0.001854	695113	3.07E-06
754911	0.001773	310485	3.07E-06
755009	0.00163	285428	3.06E-06
392407	0.00162	656215	3.06E-06
755007	0.00153	433181	3.06E-06
392448	0.001516	979773	3.06E-06
754942	0.001451	232349	3.04E-06
755011	0.001448	694636	3.04E-06
382914	0.001379	327423	3.03E-06
755010	0.001376	545225	3.03E-06
392431	0.001354	696036	3.02E-06
754870	0.001308	462679	3.02E-06
779804	0.001147	373562	3.02E-06
374915	0.001142	537670	3.01E-06
752988	0.001014	661452	3.01E-06
374887	0.000994	47911	3.01E-06
755005	0.000984	736519	3.01E-06
392406	0.000969	80723	3.00E-06
754954	0.000943	1025738	2.99E-06
374931	0.000916	857784	2.98E-06
375729	0.000876	434076	2.98E-06
754901	0.00083	977559	2.98E-06
757482	0.000828	925360	2.98E-06
374930	0.000741	418568	2.98E-06
374907	0.000675	195759	2.98E-06
392599	0.000648	713429	2.97E-06
779954	0.000647	876909	2.97E-06
754845	0.000602	285339	2.96E-06

375731	0.000598	603099	2.96E-06
754907	0.000576	53703	2.96E-06
392600	0.000569	680175	2.95E-06
755006	0.000558	1012290	2.95E-06
755013	0.000553	206001	2.95E-06
374942	0.000521	546943	2.95E-06
382915	0.000495	1051212	2.95E-06
392694	0.000455	1002749	2.95E-06
778568	0.000441	267375	2.95E-06
764316	0.000395	233853	2.95E-06
778569	0.000369	573659	2.95E-06
754944	0.00036	742154	2.94E-06
374786	0.000357	484655	2.94E-06
374963	0.000357	304690	2.94E-06
392429	0.000342	1007343	2.94E-06
375038	0.000339	474959	2.93E-06
779806	0.000335	1076774	2.93E-06
392602	0.000321	76814	2.93E-06
755012	0.000312	437294	2.93E-06
780358	0.000295	995990	2.93E-06
754894	0.000294	887609	2.92E-06
428193	0.000291	401671	2.92E-06
755016	0.000279	867027	2.92E-06
374962	0.000273	525588	2.92E-06
779810	0.000262	483052	2.92E-06
755024	0.000253	192125	2.91E-06
392446	0.000236	614823	2.91E-06
779805	0.000226	962640	2.91E-06
754836	0.00022	683291	2.91E-06
755015	0.000216	1054160	2.91E-06
764150	0.00021	193081	2.91E-06
428198	0.000206	668350	2.90E-06
778572	0.000204	684940	2.90E-06
754844	0.000197	1041666	2.90E-06
754943	0.000187	904971	2.90E-06
428192	0.000187	304142	2.90E-06
374890	0.000186	576221	2.90E-06
428197	0.000185	308152	2.90E-06
375586	0.000177	783976	2.90E-06
428194	0.000176	218098	2.90E-06

3) For TX graph

Node	Hits	Node	Pagerank
468366	0.118001	259805	3.63E-06
476960	0.066073	1200389	3.02E-06
476986	0.062455	982617	3.01E-06
476988	0.057917	114327	2.98E-06
476985	0.054796	24579	2.82E-06
476967	0.050271	114012	2.81E-06
476984	0.048178	981132	2.79E-06
476987	0.047297	992393	2.77E-06
476961	0.043748	772334	2.75E-06
468367	0.043376	82156	2.73E-06
476964	0.039549	640611	2.61E-06
476983	0.031052	972370	2.57E-06
476968	0.030067	1303097	2.57E-06
468365	0.026853	1268928	2.57E-06
476591	0.021974	445199	2.57E-06
476989	0.021893	948231	2.57E-06
476962	0.020867	980777	2.57E-06
476963	0.018276	1037360	2.57E-06
477107	0.017847	683351	2.56E-06
476998	0.015273	1227181	2.55E-06
476990	0.014746	287247	2.55E-06
476853	0.01232	98038	2.53E-06
476992	0.012088	223689	2.52E-06
476590	0.0106	535	2.52E-06
476854	0.01018	933428	2.51E-06
476997	0.008869	131052	2.50E-06
477112	0.007911	112470	2.49E-06
476996	0.00785	1383416	2.48E-06
476993	0.007393	512207	2.48E-06
476965	0.006522	1008126	2.47E-06
476592	0.005441	250887	2.46E-06
476991	0.004971	1349986	2.46E-06
477114	0.004278	110329	2.45E-06
476856	0.003802	126099	2.45E-06
476594	0.003724	237799	2.44E-06
477108	0.003292	73788	2.44E-06
476995	0.003281	1043299	2.43E-06
477115	0.002922	141407	2.43E-06
476940	0.002723	1022527	2.43E-06
476966	0.002563	674997	2.43E-06

477116	0.002466	112176	2.43E-06
476980	0.002337	1260033	2.42E-06
476584	0.001683	1348698	2.42E-06
477113	0.001436	823684	2.41E-06
476939	0.001426	632344	2.41E-06
469472	0.001176	19183	2.41E-06
477001	0.001168	566908	2.41E-06
477117	0.001039	348057	2.40E-06
476852	0.00092	1224332	2.40E-06
476846	0.000812	993089	2.40E-06
476907	0.000766	1046025	2.38E-06
468780	0.000756	1261194	2.38E-06
476957	0.000714	209307	2.38E-06
476982	0.000704	1065557	2.38E-06
476969	0.00065	972670	2.38E-06
476593	0.000635	265232	2.37E-06
476994	0.000574	657992	2.37E-06
476941	0.000518	133913	2.37E-06
476832	0.000506	723794	2.37E-06
477143	0.000379	982061	2.37E-06
476582	0.000378	1075760	2.37E-06
469473	0.00037	647387	2.37E-06
477121	0.000357	329267	2.36E-06
468779	0.000339	1176518	2.36E-06
133913	0.000278	1338526	2.36E-06
476999	0.000261	882481	2.36E-06
476979	0.000259	417087	2.36E-06
476981	0.000236	1249727	2.35E-06
476970	0.000229	51172	2.35E-06
468781	0.000213	821767	2.35E-06
477126	0.000206	367276	2.35E-06
477007	0.000199	385597	2.34E-06
476976	0.000174	22669	2.34E-06
476977	0.000154	651906	2.34E-06
476831	0.000153	327268	2.34E-06
134887	0.000149	1127635	2.34E-06
476833	0.000146	288584	2.34E-06
477000	0.000145	1322231	2.34E-06
134680	0.000145	840340	2.33E-06
476942	0.000137	825725	2.33E-06
134679	0.000136	1159354	2.33E-06
476958	0.000135	90327	2.33E-06
476950	0.000127	801175	2.33E-06

134678	0.000125	89975	2.33E-06
477122	0.000123	238422	2.33E-06
480860	0.00012	853216	2.32E-06
134874	0.000117	793620	2.32E-06
134866	0.000111	958284	2.32E-06
133914	9.67E-05	1143756	2.32E-06
134865	9.34E-05	544127	2.32E-06
476583	8.80E-05	418108	2.32E-06
476882	8.04E-05	726501	2.32E-06
134878	7.17E-05	925294	2.31E-06
469471	7.06E-05	95556	2.31E-06
476975	7.05E-05	911390	2.31E-06
476978	6.94E-05	1243916	2.31E-06
479392	6.88E-05	569395	2.31E-06
476845	6.57E-05	1172237	2.31E-06
477011	6.40E-05	825992	2.31E-06
476908	5.57E-05	697206	2.31E-06

C)

Yes the results differ significantly, because the hits algorithm consists takes into consideration both authorities (in links) and hubs (out links). For instance, if initially all the nodes are given the same hub scores, the authority values for the nodes with most neighbors is highest. While computing the hub scores again, nodes whose neighbors have higher authority scores have better hub scores for next iteration and so on. Thus the algorithm tends to give high scores to group of nodes that are highly connected among each other (degree centrality) but may not be most critical in to the entire graph.

Page rank on the other hand considers the probability that you would arrive to a node in a random walk. It only considers the in links to the graph. Thus a node may not have higher degree, but its connection to nodes with higher scores will increase the score of the node. It is more analogues to Eigen vector centrality.

2)

a) The following methods can be applied to develop a keyword list:

1) Look at the past data when there was actually a flu and see the most frequently searched terms from the search data. We can prepare a list first that is linguistically correlated to flu (eg: medicines, fever, cough etc.) and find the common words in the list and the frequent search data during the flu season. It may be necessary to collaborate with a medical agency such as CDC so that we get data of the actual reported cases for verification. We can use the data obtained from CDC and find words whose trends have high correlation with that data.

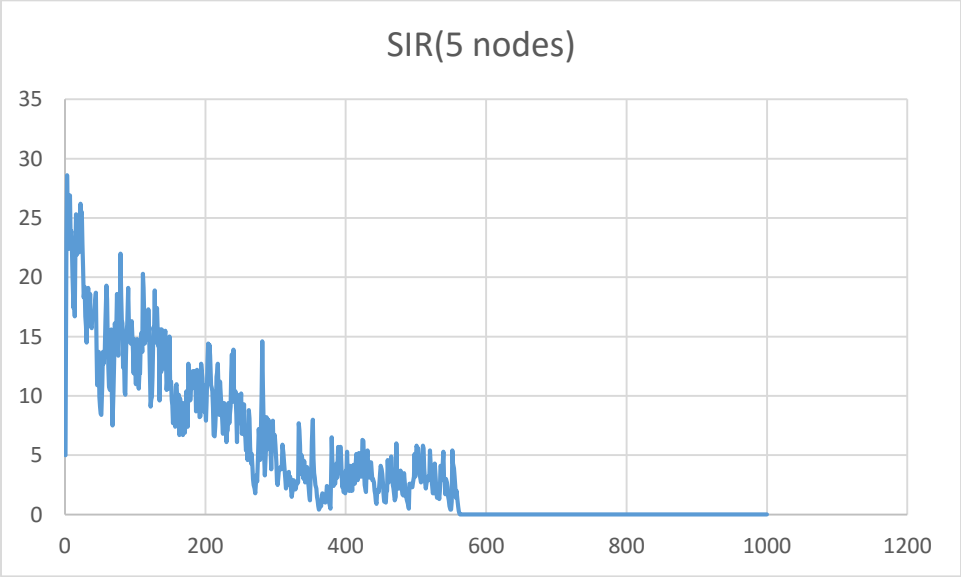
2) Simply discard the queries that are uncorrelated to the search data. Same analysis can be applied to social media data. Then we can assign a score to the filtered list based on the frequency of the search terms.

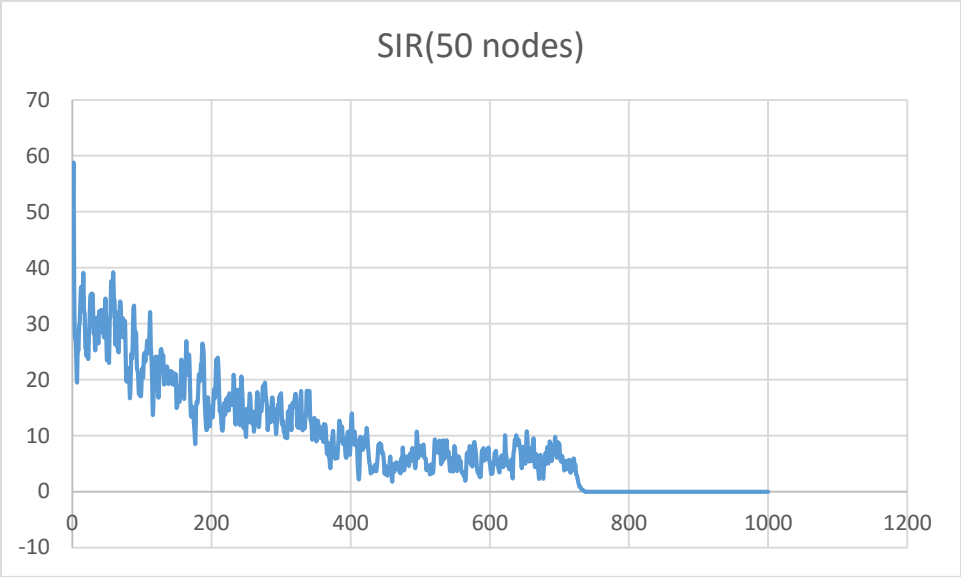
b) Yes we can modify the list further based on geography. We can add to our filtered list all "synonyms" commonly used in a particular region. For instance as suggested in the question, add soda to the list if 'Pepsi' exists in the list. Similar technique can be applied as suggested in part A. Find the keywords that have similar trends as the flu patterns, but on search queries in a particular region

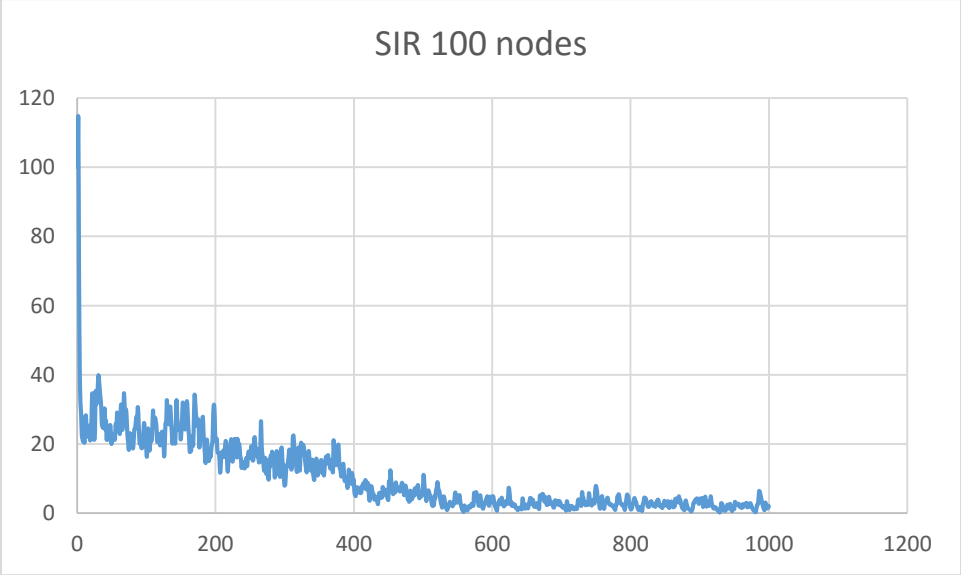
c) The same thing can be done for a Spanish. Find a list of words commonly used in Spanish related to the above list and add them to the list. We can assign higher scores to words commonly used locally.

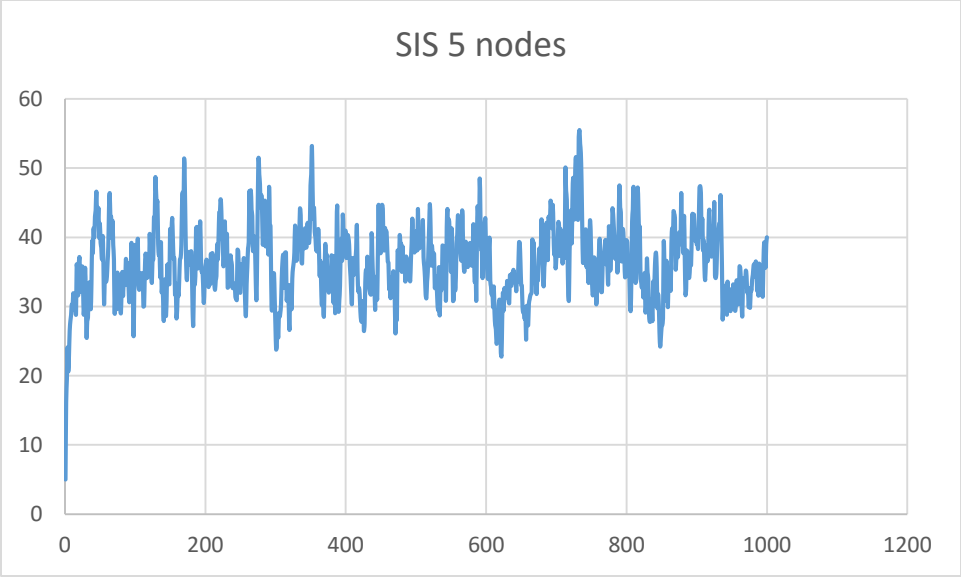
4) For Sir Model

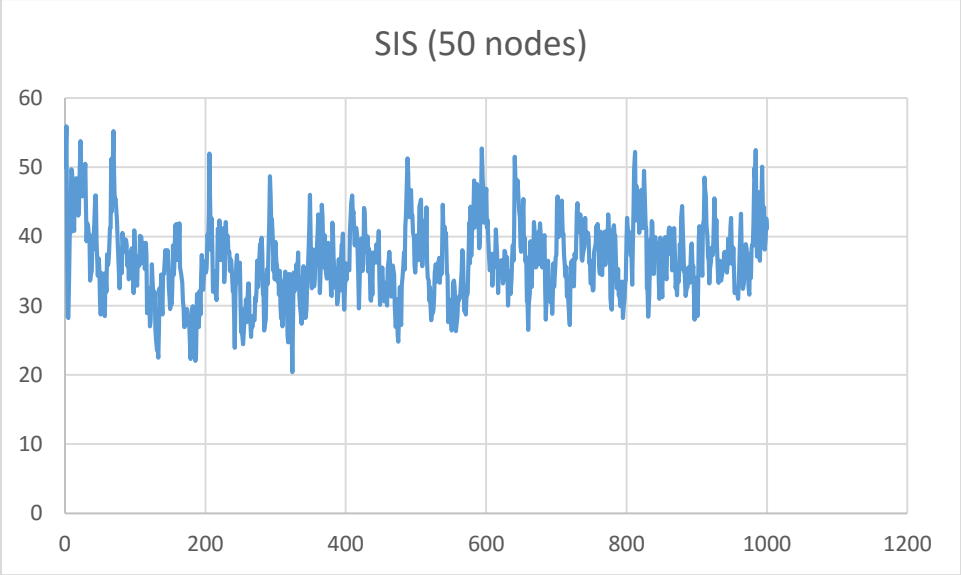
1)

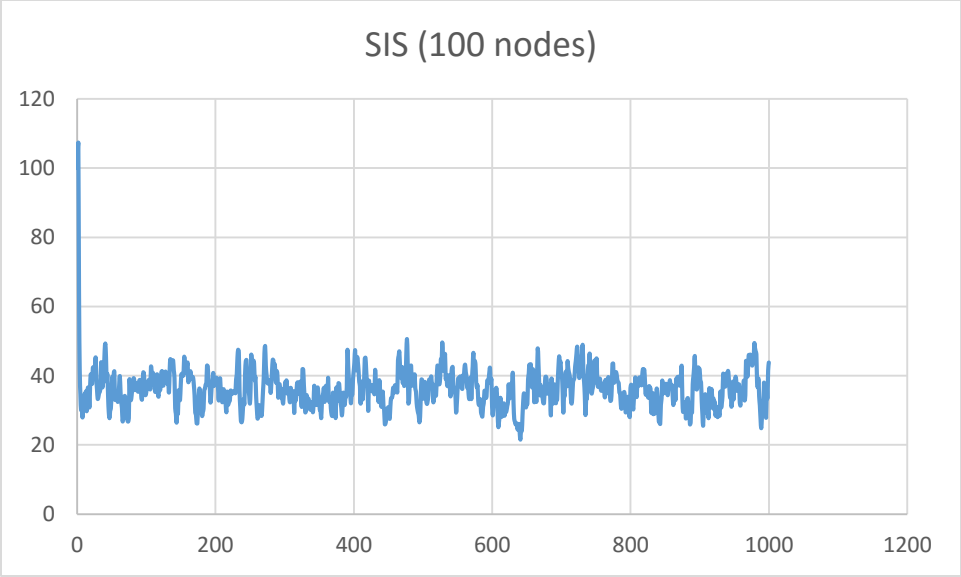


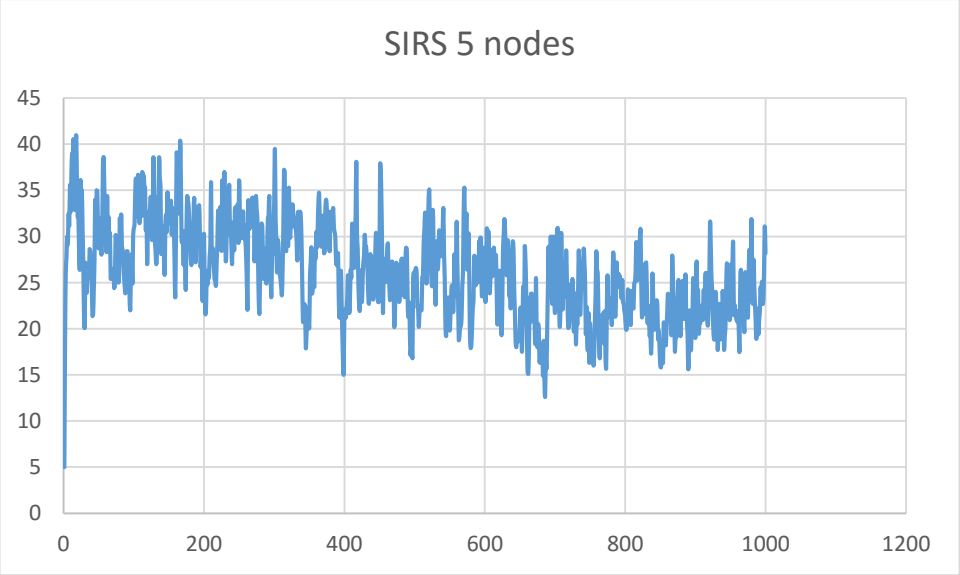

SIR(50 nodes)
70
60
50
40
30
20
10
0
-10
0
200
400
600
800
1000
1200

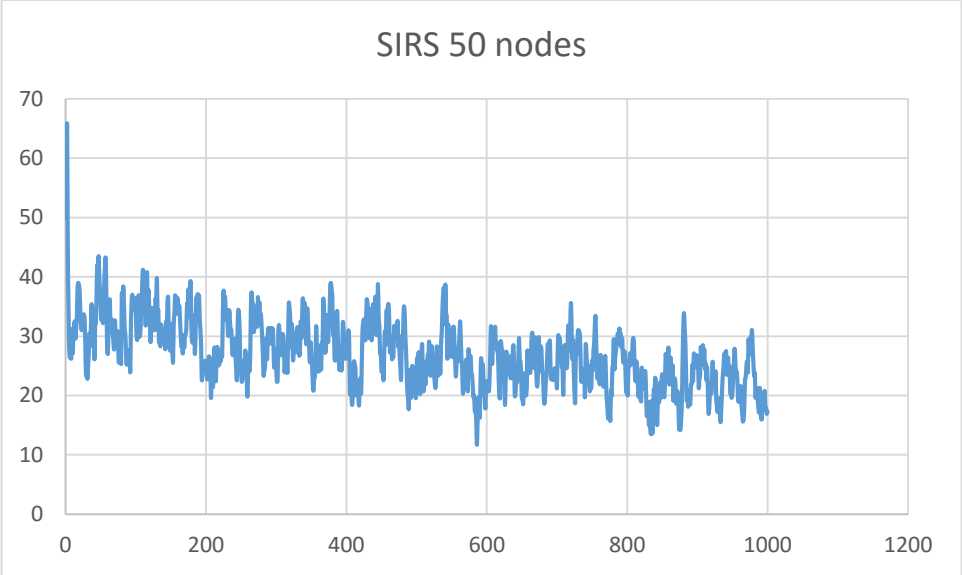

SIR 100 nodes
120
100
80
60
40
20
0
0
200
400
600
800
1000
1200

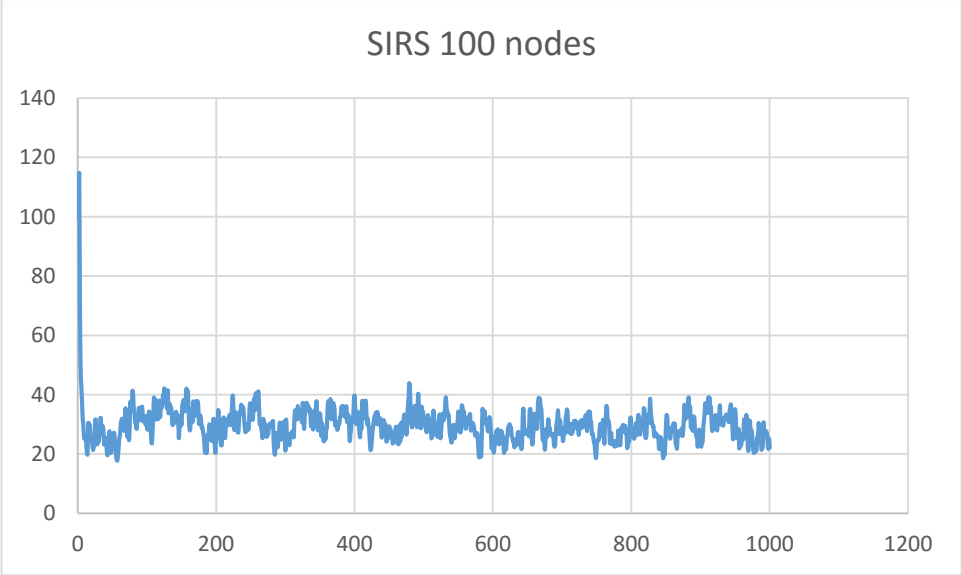

SIS 5 nodes
60
50
40
30
20
10
0
0
200
400
600
800
1000
1200


SIS (50 nodes)
60
50
40
30
20
10
0
0
200
400
600
800
1000
1200


SIS (100 nodes)
120
100
80
60
40
20
0
0
200
400
600
800
1000
1200

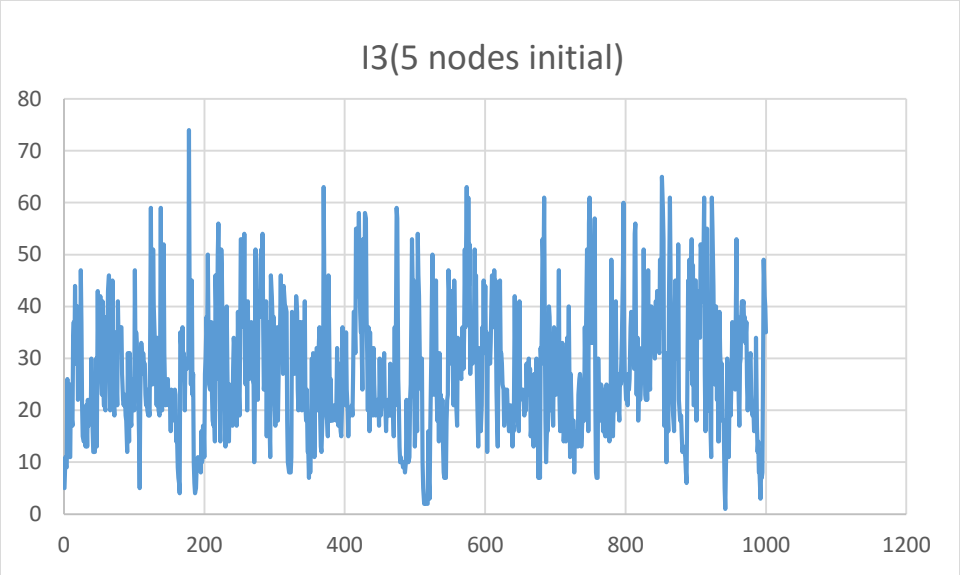

SIRS 5 nodes
45
40
35
30
25
20
15
10
5
0
0
200
400
600
800
1000
1200

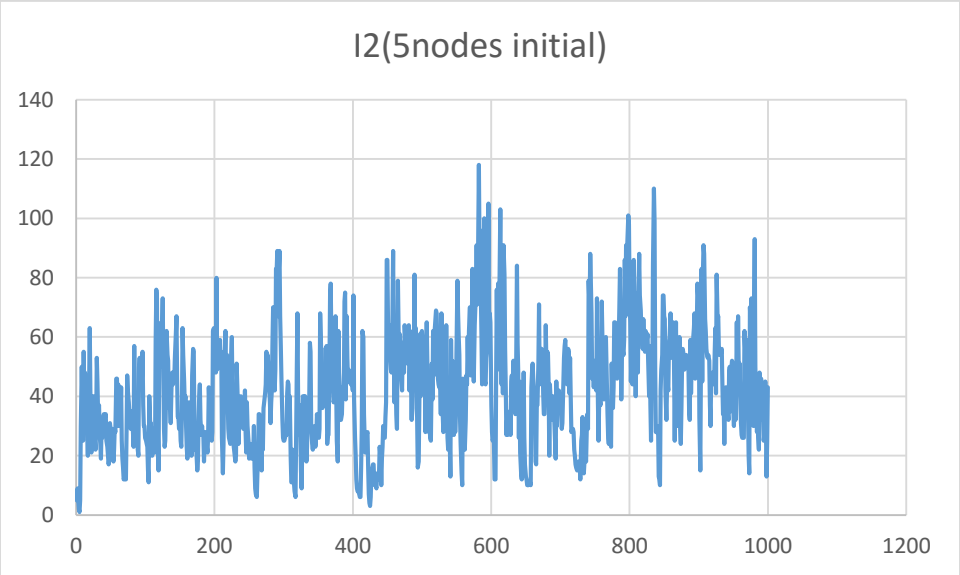


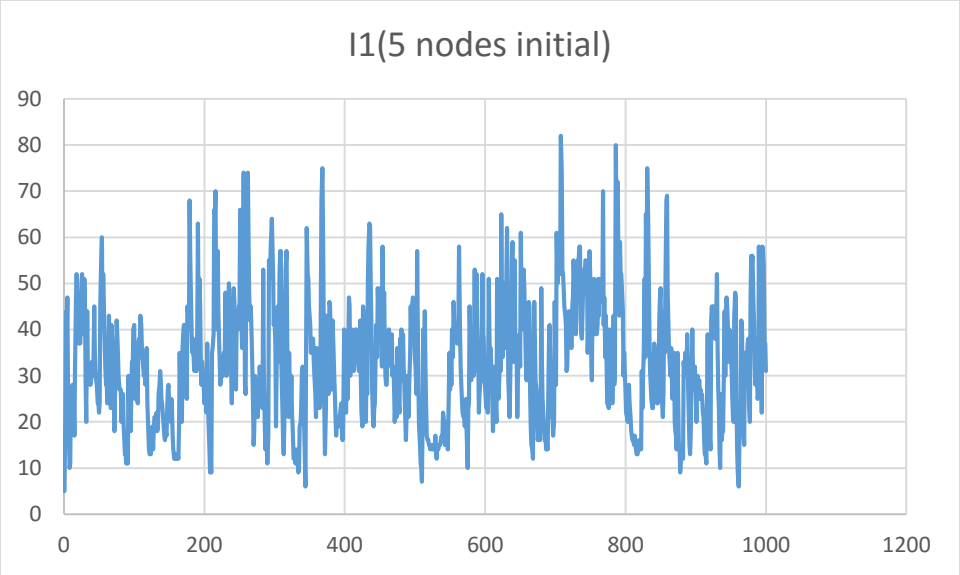


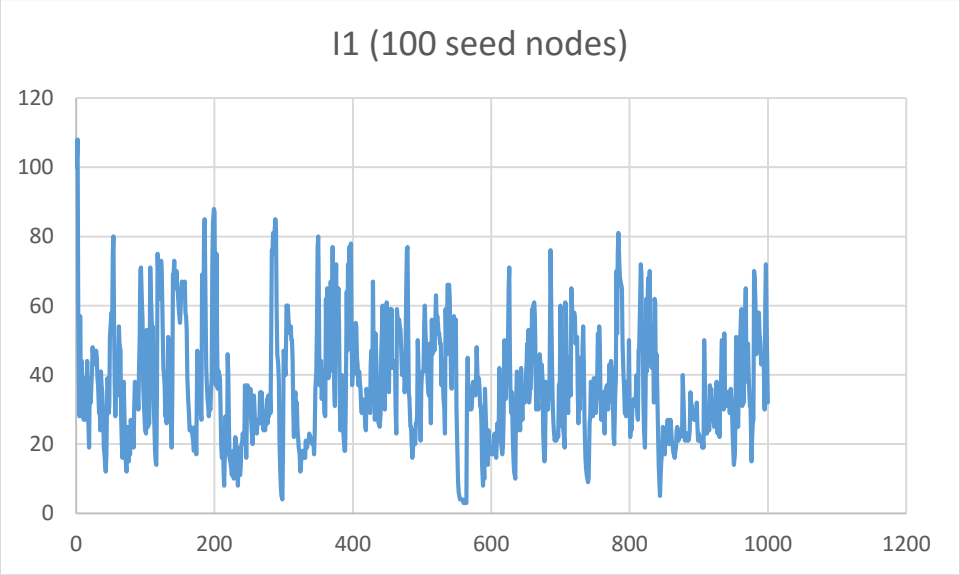
The Epicurve for SIR falls to 0 because the recovery rate is higher than the infection rate. For SIS, the number of infected nodes hovers around a particular value but does not go to zero, while for SIRS the number of infected people also fluctuates around a particular value, but the value is less than that of SIS model. All the simulations are done are averaged on 10 (running in 10 iterations took a long time, couldn't perform for 1000 iterations as suggested on piazza) iterations.

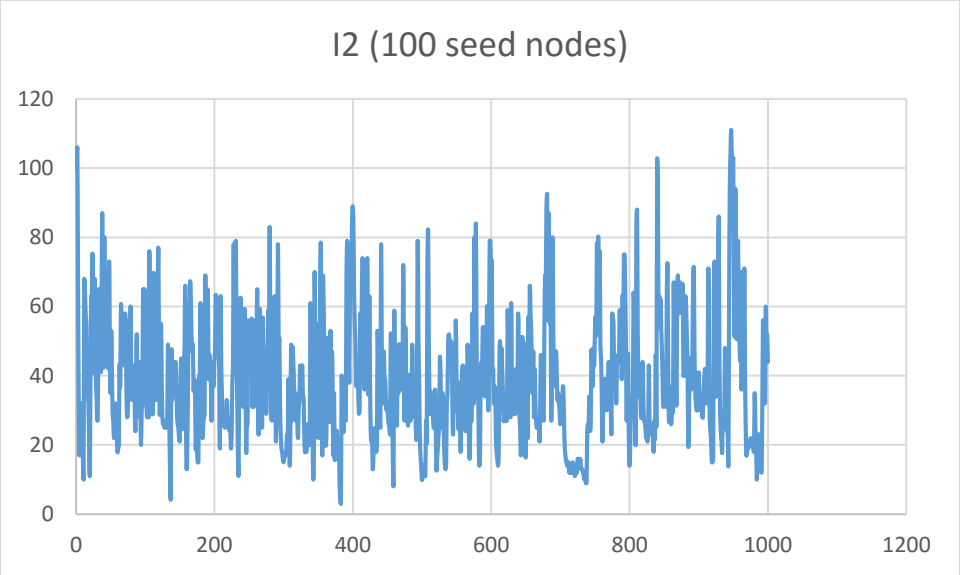
b)

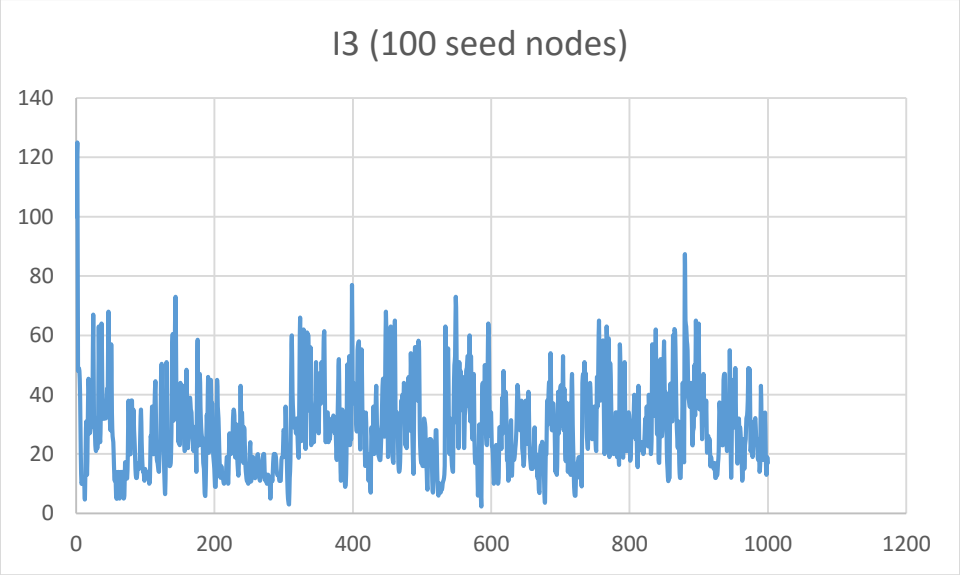

I3(5 nodes initial)
80
70
60
50
40
30
20
10
0
0
200
400
600
800
1000
1200


I2(5nodes initial)
140
120
100
80
60
40
20
0
0
200
400
600
800
1000
1200


I1(5 nodes initial)
90
80
70
60
50
40
30
20
10
0
0
200
400
600
800
1000
1200


I1 (100 seed nodes)
120
100
80
60
40
20
0
0
200
400
600
800
1000
1200





Yes all virus exist in the system at the end. The infection with higher recovery rate will die down quickly, whereas the infection with higher beta (infection rate) will dominate the rest of the infections. The behavior is apparent in the simulation with 100 seed nodes. I1 (0.3, 0.5),I2(0.4,0.6) have same number of infected nodes as I2 has higher beta but gamma is also higher which balances it out.

I3 has lower number of infected nodes because of lower beta and higher gamma.