

Visualization of CRAN Package Dependency Network

Group 3: X Zhang, X Chen, W Qiwen, N Xiao

2010 PKU Vis Summer School

August 18, 2010

Slides now available at:

<http://vis.pku.edu.cn/wiki/>

Public Courses \Rightarrow Summer School s10 \Rightarrow Group 3

Outline

- 1 Why
- 2 How
- 3 What we got
- 4 Reference
- 5 Appendix

What is **CRAN**?



Comprehensive R Archive Network

CTAN: Comprehensive TeX Archive Network

CPAN: Comprehensive Perl Archive Network

Why needs to be VISUALIZED?

- 1 Real world data beside us
- 2 Strengthen the open source community
- 3 Has never been VISUALIZED before

Data Extracted from CRAN:
Raw Text.

After Pre-Processing:
1805 Nodes, **4020** Edges.
Now available at Vis wiki.

- Tools:

- 1 Pajek (1996)

- 2 Gephi (2008)

- 3 Network Manipulator (McGuffin, 2009)

- Why not R

- 1 Basic SNA
- 2 “Big Picture”
- 3 Manual Analysis

Basic SNA

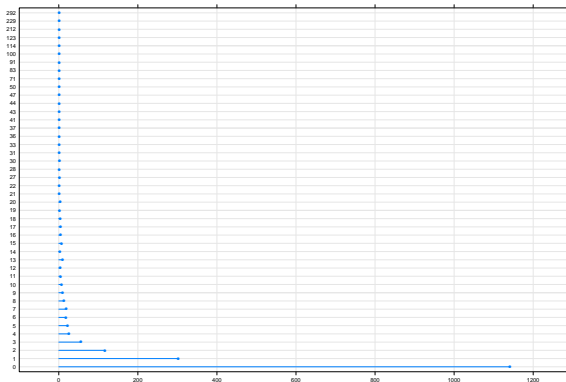


Figure: Output Degree of the Network

Basic SNA

Diameter:

`stats` \Rightarrow `tossm`

Distance: 5.

Basic SNA

Clustering Coefficients:

$$CC(i) = \frac{2E_1(i)}{(k_i(k_i - 1))}$$

$$CC = \frac{\sum_{i=1}^N CC(i)}{N}$$

For this network: $CC = 0.0697$

Hubs and Authority

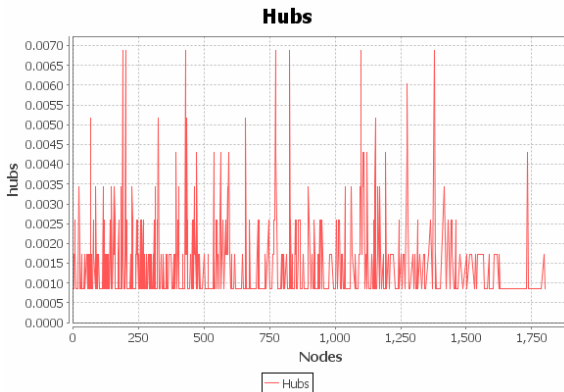


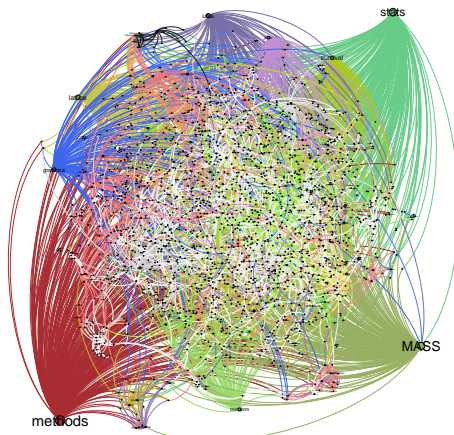
Figure: Hubs

Hubs and Authority



Figure: Authority

“Big Picture”



“Big Picture”

Gephi project and PDF(Vector) Version also available at Vis wiki.

Force Directed Layout

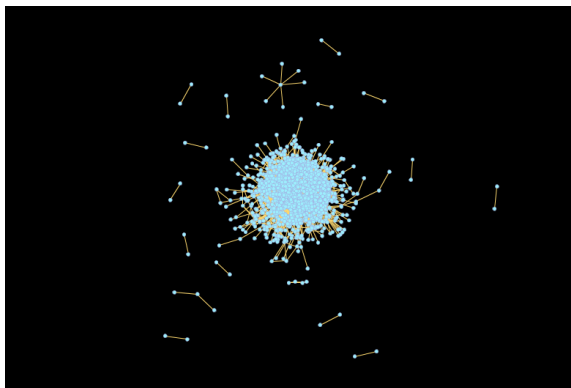


Figure: Force Directed Graph by GUESS

Manual Analysis: Most Important Packages

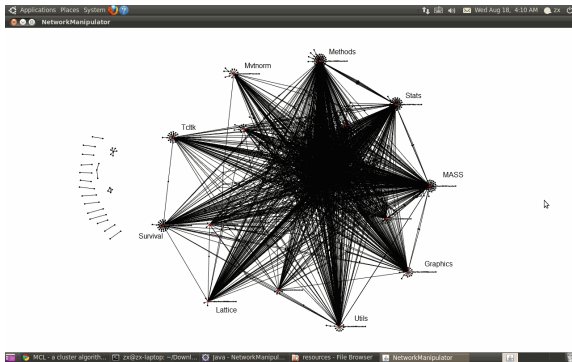


Figure: Manual Analysis: Most Important Packages (1)

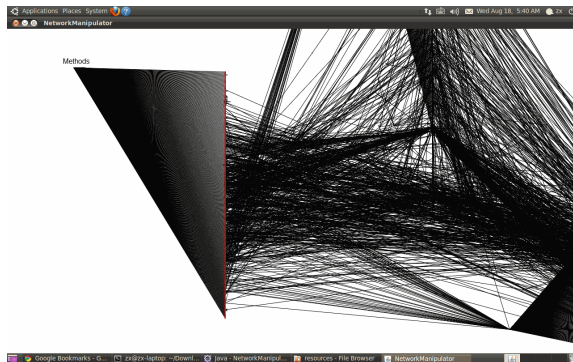


Figure: Manual Analysis: Most Important Packages (2)

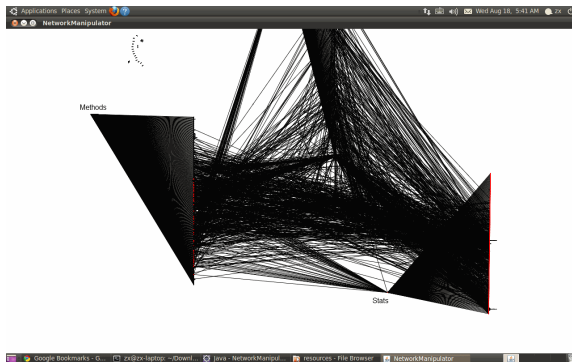


Figure: Manual Analysis: Most Important Packages (3)

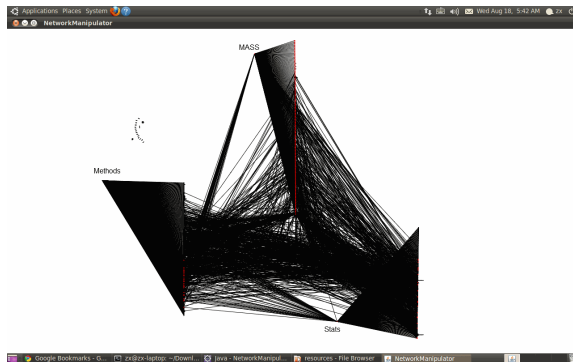


Figure: Manual Analysis: Most Important Packages (4)

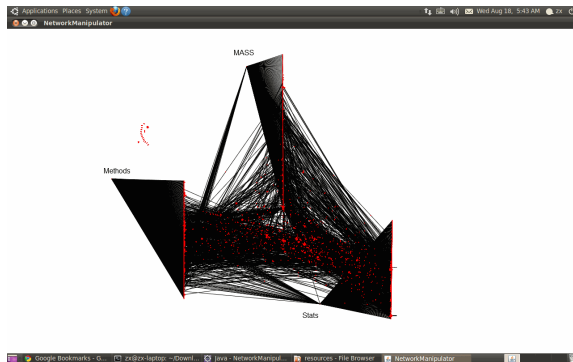


Figure: Manual Analysis: Most Important Packages (5)

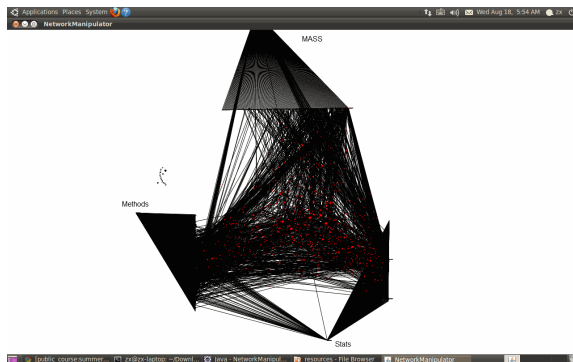


Figure: Manual Analysis: Most Important Packages (6)

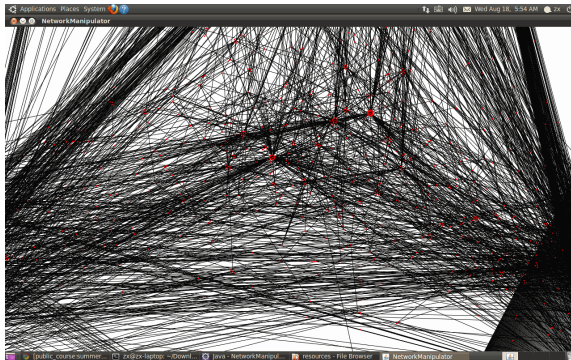


Figure: Manual Analysis: Most Important Packages (7)

1	18	19	37	112	117	165	166	167	168	188	191	192	193	200	205	211	250	252	256	315	316	363	367	392	404	405	454	525	537	538		
	549	570	571	577	579	582	589	590	592	593	594	595	596	599	602	615	619	669	672	673	675	680	712	722	729	742	751	752	753			
	754	755	759	760	767	768	773	776	811	820	821	823	826	846	851	852	854	855	878	879	881	897	898	911	918	924	928	939	940	952		
	986	988	989	990	1013	1014	1015	1016	1017	1020	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112		
	1113	1114	1115	1117	1128	1129	1130	1138	1140	1141	1154	1164	1165	1167	1168	1180	1183	1202	1206	1213	1221	1229	1236	1242	1249	1255	1256	1273	1274	1283		
	1290	1293	1304	1305	1321	1327	1338	1386	1411	1412	1422	1429	1438	1440	1445	1482	1485	1486	1495	1507	1511	1512	1521	1522	1533	1544	1545	1561	1567	1573		
	1582	1583	1600	1619	1668	1672	1723	1726	1735	1740	1744	1751	1758	1767	1790																	
2	1	2	4	7	48	80	81	113	115	116	130	148	149	158	161	176	204	206	241	278	344	345	353	368	395	396	428	441	446	447	448	452
	453	466	489	490	529	530	563	621	666	676	678	716	717	718	825	847	865	872	882	904	920	921	922	930	934	944	945	948	959	960		
	969	991	996	1024	1046	1047	1052	1061	1063	1074	1075	1076	1077	1079	1083	1085	1112	1127	1131	1135	1155	1158	1163	1198	1233	1248	1271	1281	1310	1322		
	1337	1352	1376	1395	1409	1410	1419	1426	1427	1428	1437	1455	1456	1457	1458	1465	1474	1487	1489	1494	1506	1519	1536	1542	1546	1551	1557	1558	1584	1633		
	1636	1646	1654	1662	1666	1670	1701	1713	1750	1779	1782	1796																				
3	22	23	24	77	78	83	194	219	257	262	264	311	329	358	359	399	429	430	439	451	464	471	473	478	498	502	503	521	573	578		
	581	663	681	683	685	728	734	783	785	794	813	814	816	876	890	895	905	906	953	963	997	1051	1058	1062	1139	1142	1166	1176	1211	1228		
	1240	1246	1297	1306	1347	1351	1365	1366	1368	1371	1407	1432	1435	1451	1470	1483	1484	1492	1497	1498	1501	1503	1508	1547	1575	1610	1611	1651	1697	1699		
	1791	1801																														
4	8	14	141	142	174	175	314	386	394	412	457	458	488	496	533	638	651	652	696	860	867	931	956	958	966	970	971	972	973	1069		
	1133	1153	1210	1219	1277	1296	1299	1349	1377	1394	1423	1460	1490	1509	1515	1525	1530	1553	1596	1607	1609	1656	1671	1714	1715	1716	1736	1775				
5	20	21	49	61	62	64	65	66	107	233	234	286	310	362	393	440	622	692	739	842	859	877	885	886	933	975	1006	1181	1192	1235		
	1266	1311	1341	1342	1361	1363	1378	1398	1406	1410	1448	1471	1472	1652	1702	1708	1725	1728	1741													
6	30	103	106	120	214	273	354	355	500	562	572	614	627	639	703	746	747	822	1036	1037	1065	1132	1197	1265	1326	1385	1389	1480	1574	1594		
	1721	1722	1770																													
7	29	133	384	385	510	617	667	674	690	707	866	896	901	983	1018	1026	1027	1028	1029	1066	1292	1394	1590	1622	1661							
8	67	172	246	374	378	494	568	799	856	861	862	985	1000	1073	1250	1253	1254	1272	1318	1320	1354	1369	1439	1653	1787							
9	93	94	95	566	580	605	606	616	668	684	731	789	1200	1217	1222	1538	1585	1616	1642	1771												
10	164	432	482	611	708	709	723	774	775	777	849	850	981	1019	1143	1572	1623	1693	1694	1698												
11	118	125	390	391	567	574	583	584	585	796	819	837	838	962	1238	1576	1588	1637	1783													
12	260	765	766	769	801	802	803	1012	1247	1295	1315	1360	1513	1523	1526	1527	1529	1608	1753													
13	69	160	173	189	501	719	974	1067	1071	1082	1311	1400	1579	1649	1657	1674	1737	1784														
14	5	9	114	371	372	445	526	1044	1150	1152	1570	1648	1727	1730	1733	1794																

Figure: Clustering Result by MCL

Future Jobs

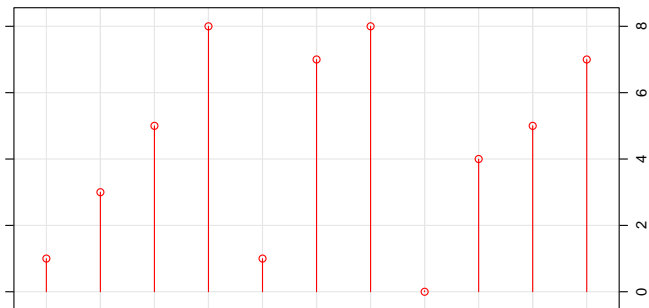
1-mode \Rightarrow 2-mode

Authors + Packages

Reference

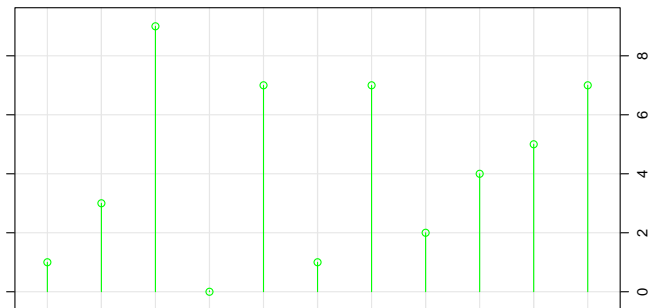
R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.

The Namecards



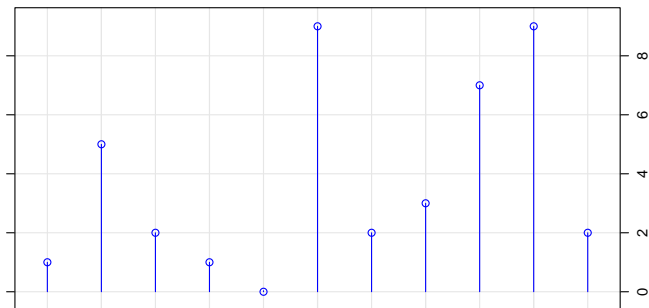
Zhang Xin

The Namecards



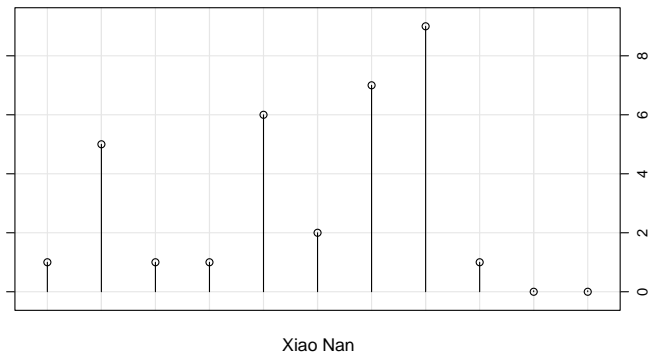
Chen Xuan

The Namecards



Qiwen Wang

The Namecards



Questions?

Thanks