Measuring UK Crime Networks

Giles Oatley and Tom Crick
tcrick@cardiffmet.ac.uk

Department of Computing & Information Systems
Cardiff Metropolitan University, UK
http://drtomcrick.com

19 August 2014

Cardiff Metropolitan University | Prifysgol Metropolitan Caerdydd

Abstract

This paper describes the output of a study to tackle the problem of gang-related crime in the UK; we present the intelligence and routinely gathered data available to a UK regional police force, and describe an initial social network analysis of gangs in the Greater Manchester area of the UK between 2000-2006.

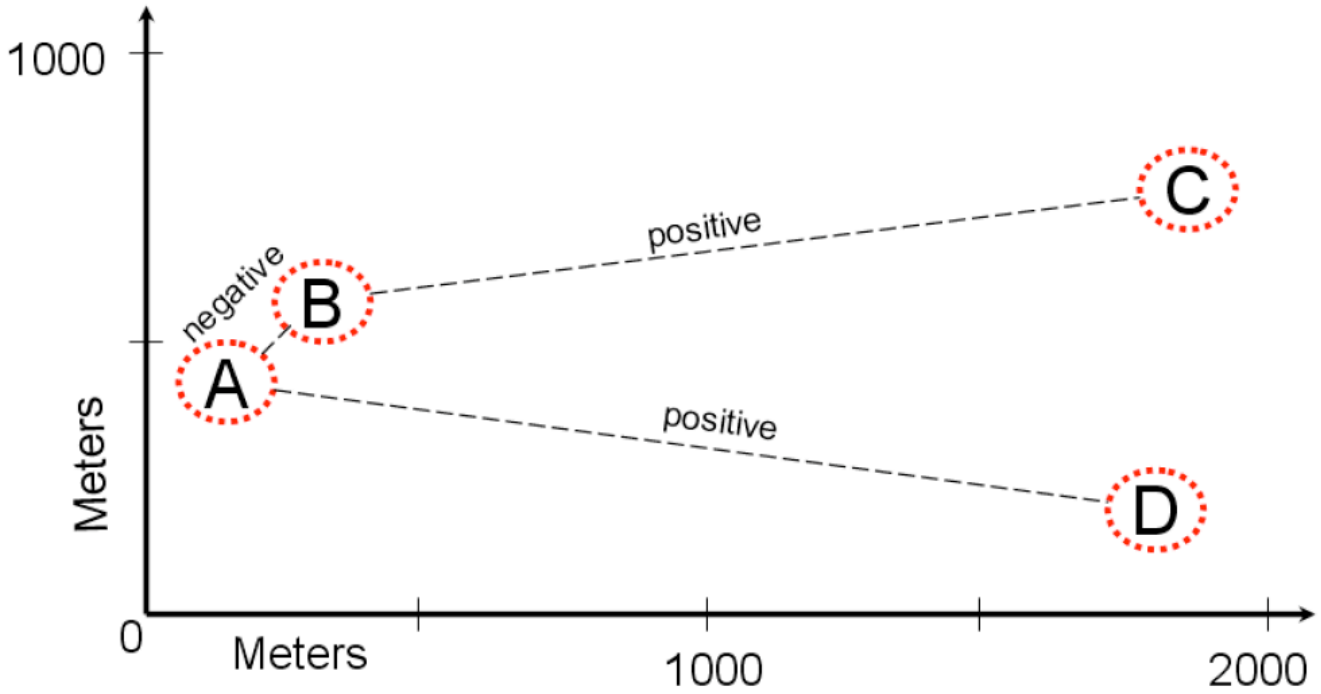
By applying social network analysis techniques, we attempt to detect the birth of two new gangs based on local features (modularity, cliques) and global features (clustering coefficient). Thus for the future, identifying the changes in these can help us identify the possible birth of new gangs (sub-networks) in the social system.

Furthermore, we study the dynamics of these networks globally and locally, and have identified the global characteristics that tell us that they are not random graphs – they are small world graphs – implying that the formation of gangs is not a random event. However, we are not yet able to conclude anything significant about scale-free characteristics due to insufficient sample size.

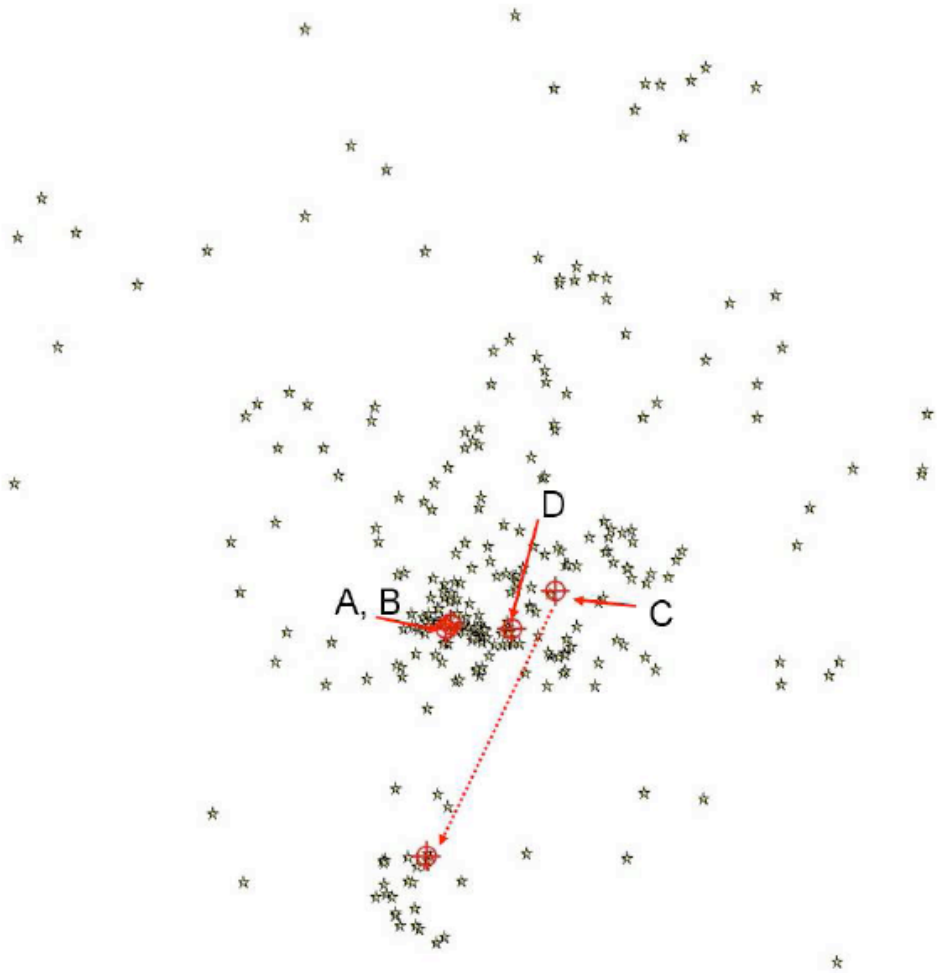
Gang Formation

Gang label	Gang Name	Formation
A	Gooch	1990s
B	Doddington/Pepperhill	1990s
C	Longsight Crew	c.2001
D	Rusholme Crew Gangsters	c.2004

Table I. GANG NAMES AND APPROXIMATE DATES OF FORMATION.



Gang Formation

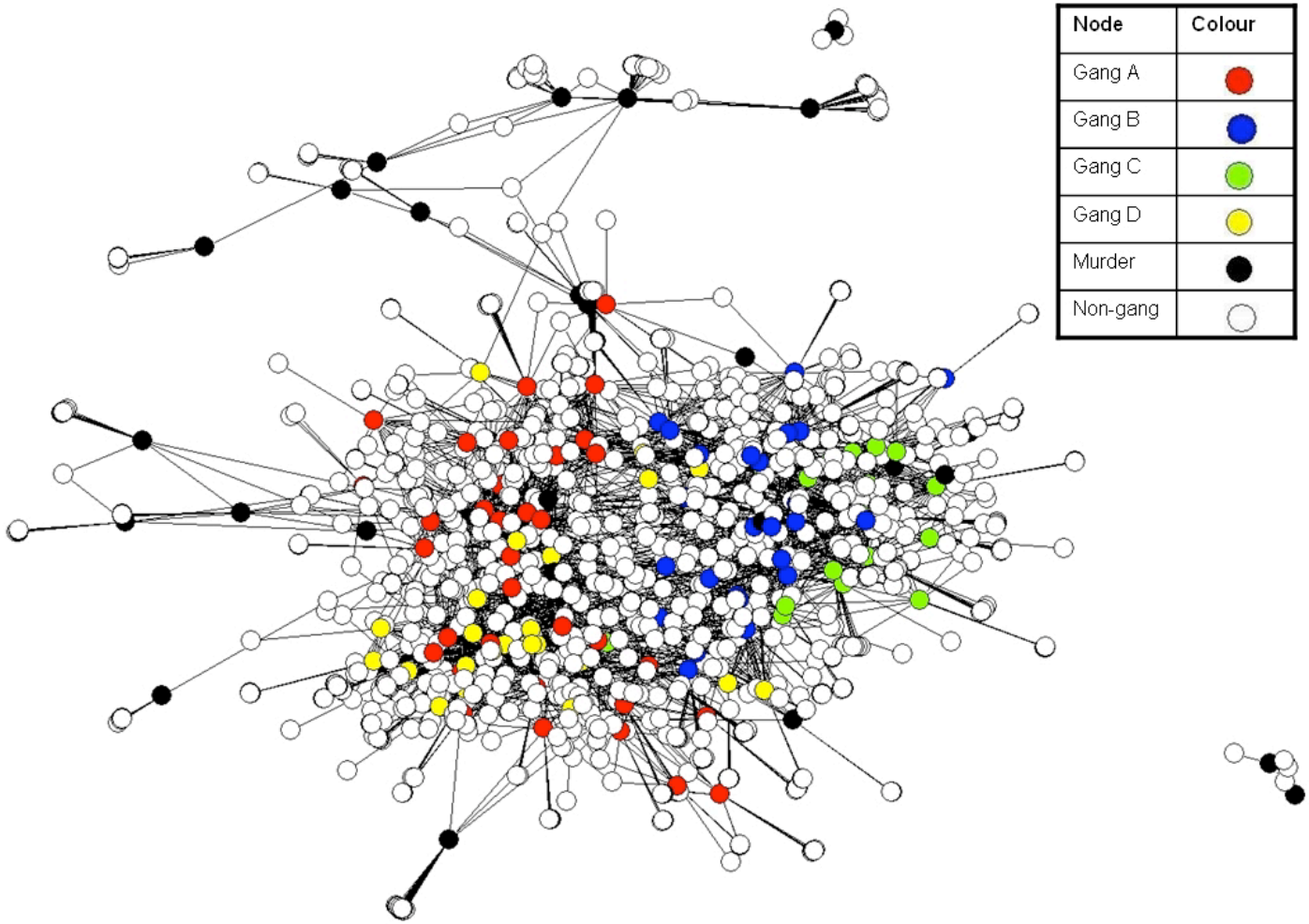


Gangs in South Manchester, UK

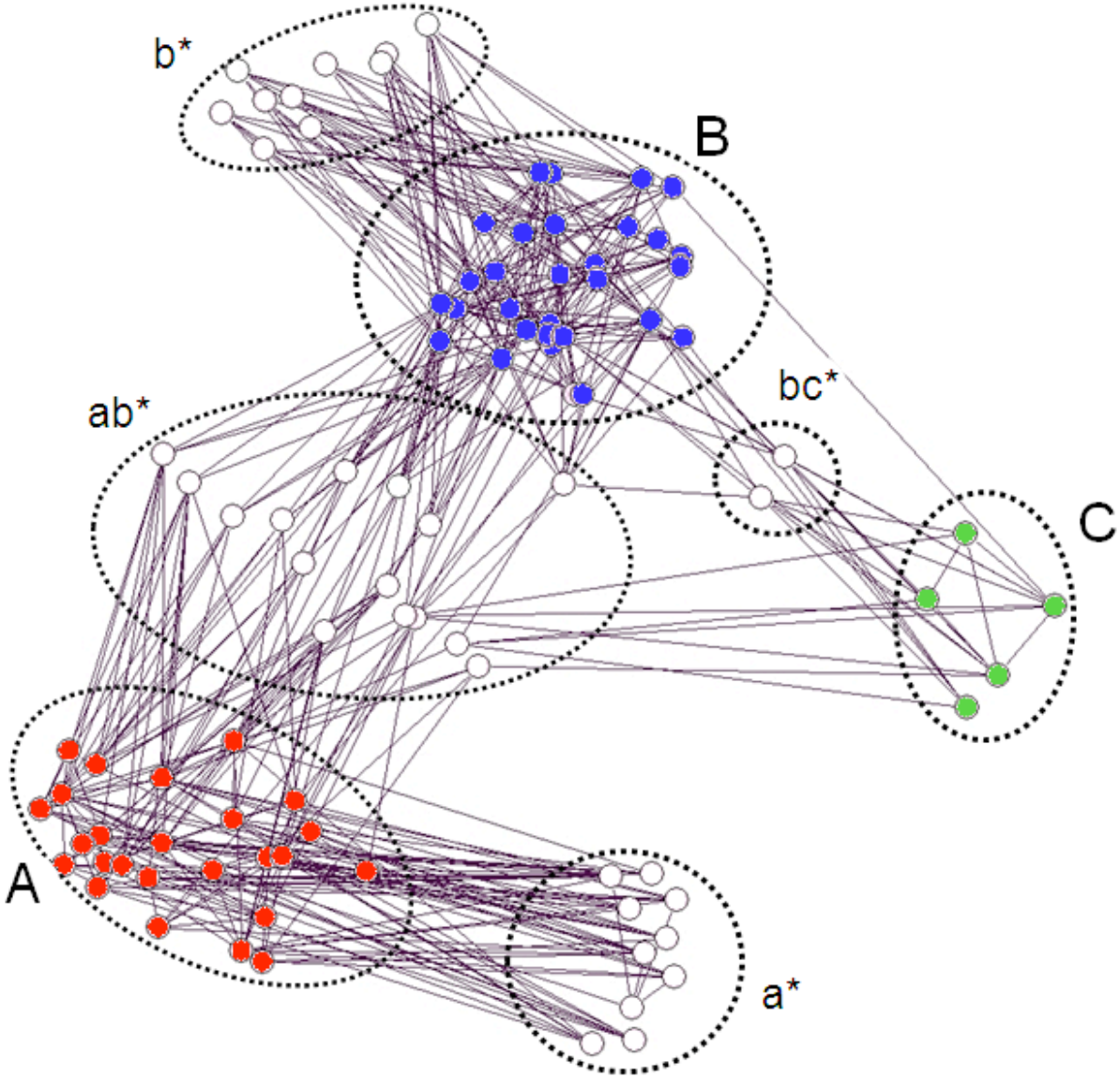
- All serious crimes: *murder, attempted murder, manslaughter, kidnapping, serious wounding, firearms offences*.
- Gang C has moved into an additional location with drug selling.
- 38% (n=162) of all serious crimes occurring within 1 km radius (of gang locations).
- 63% of all serious crimes occur within 2 km.
- 53% (n=9) of murders are within 3 km
- 38% (n=34) of attempted murders are within 1 km, 63% within 2 km
- 33% (n=17) of serious woundings are within 1 km, 48% are within 2 km.

Magnet Category	Relationship	Frequency/Percentage	Reason Reported Examples
Crime related	1. Accomplice	502, 10.7%	(i.) Arrested Together (ii.) Believed To Be Dealing Drugs Together (iii.) X's Sister Is Y's Girlfriend
	2. Charged with	45, 1.0%	(i.) Charged Together Murder (ii.) Arrested Together
Familial	3. Brother	65, 1.4%	(i.) Believed To Be Half-Brothers
	4. Child	23, 0.5%	(i.) Father & Son (ii.) Admitted Above Named Is His Dad
	5. Parent	20, 0.4%	(i.) Mother & Son
	6. Relative	173, 3.7%	(i.) Cousins (ii.) X States Y Is His Uncle
	7. Sister	18, 0.4%	(i.) Brother And Sister (ii.) Stated They Are Brothers
	8. Spouse	2, 0.0%	(i.) Arrested Together Handling
Friendships	9. Cohabitant	5, 0.1%	(i.) Possibly Living Together At Anon Street
	10. Friend	1409, 30.0%	(i.) Stop Checked Together In car (ii.) Attended Club Together (iii.) Seen Together
	11. Girlfriend	61, 1.3%	(i.) Have Child Together
	12. Boyfriend	10, 0.2%	(i.) Girlfriend/Boyfriend
Other	13. Other	2364, 50.3%	(i.) Ex-Boyfriend Of The Above Named (ii.) R Claimed E Stabbed Him (iii.) C Intends Killing A/N Re Murder Of Bros (iv.) Tog At Nightclub, Oldham (v.) Seen Together (vi.) Attended Murder Trial (vii.) Arrested Together In Anon (viii.) D's Number In C's Mobile (ix.) Seen Together At Moss Side Festival

Table II. EXAMPLES OF THE 'ASSOCIATES' DATA. THIS DATA IS USED TO CREATE THE SOCIAL NETWORKS. GANG MEMBERSHIP COMPRISES A MIX OF SAME-AGE LOCAL FRIENDSHIP GROUPS, BLOOD RELATIVES AND RECRUITS: UK HOME OFFICE REPORT [27].



Network in 2006.



Link reduction, showing Gangs A and B and emergence of Gang C (for 2002). This also illustrates the large amount of non-gang members who are associated with individual gangs (*a**, *b**) or who are intermediaries (*ab**, *bc**).

Small-World

- A small-world network has both local connectivity and global reach, and is a simple connected graph G exhibiting two properties:
  1. Small characteristic path length: the presence of short-cut connections between some vertices results in a small characteristic path length $L(G)$.
  2. Large clustering coefficient: each vertex of G is linked to a relatively well-connected set of neighbouring vertices, resulting in a large value for the clustering coefficient $C(G)$.

Small-World

- A small-world network has both local connectivity and global reach, and is a simple connected graph G exhibiting two properties:
  1. Small characteristic path length: the presence of short-cut connections between some vertices results in a small characteristic path length $L(G)$.
  2. Large clustering coefficient: each vertex of G is linked to a relatively well-connected set of neighbouring vertices, resulting in a large value for the clustering coefficient $C(G)$.
- To determine whether our network is a random one or is small-world, we can test whether or not it has exponential k-connectivity distribution.

Small-World

- A small-world network has both local connectivity and global reach, and is a simple connected graph G exhibiting two properties:
  1. Small characteristic path length: the presence of short-cut connections between some vertices results in a small characteristic path length $L(G)$.
  2. Large clustering coefficient: each vertex of G is linked to a relatively well-connected set of neighbouring vertices, resulting in a large value for the clustering coefficient $C(G)$.
- To determine whether our network is a random one or is small-world, we can test whether or not it has exponential k-connectivity distribution.
- **We do not observe this in the data**, however, we do see large clustering coefficients, and the average path lengths are always less than $log(n)$.
- **Based upon these two criteria we can still conclude that our networks have small-world characteristics.**

Scale-Free

- Plotting the clustering coefficient as a function of the number of nodes n, should follow the power-law distribution for scale-free networks, with the clustering coefficient being roughly four times larger than random networks.
- The value of the clustering coefficient for a random networks will be $1/n$.

Scale-Free

- Plotting the clustering coefficient as a function of the number of nodes n, should follow the power-law distribution for scale-free networks, with the clustering coefficient being roughly four times larger than random networks.
- The value of the clustering coefficient for a random networks will be $1/n$.
- We compare the values of $4/n$ against CC.
- As the cumulative links increase from 2000 to 2006, the value of CC generally increases (with the number of nodes n) and is always significantly higher than the values of $4/n$.
- Each of the gang values for CC are also significantly higher than would be expected in a random network.

Scale-Free cont.

- The diameter of the network (longest path length) should be approximately *log(log(n))* for scale-free networks.
  - **In both cases (for the gangs and the years) the real values are significantly higher than would be expected for a scale-free network.**

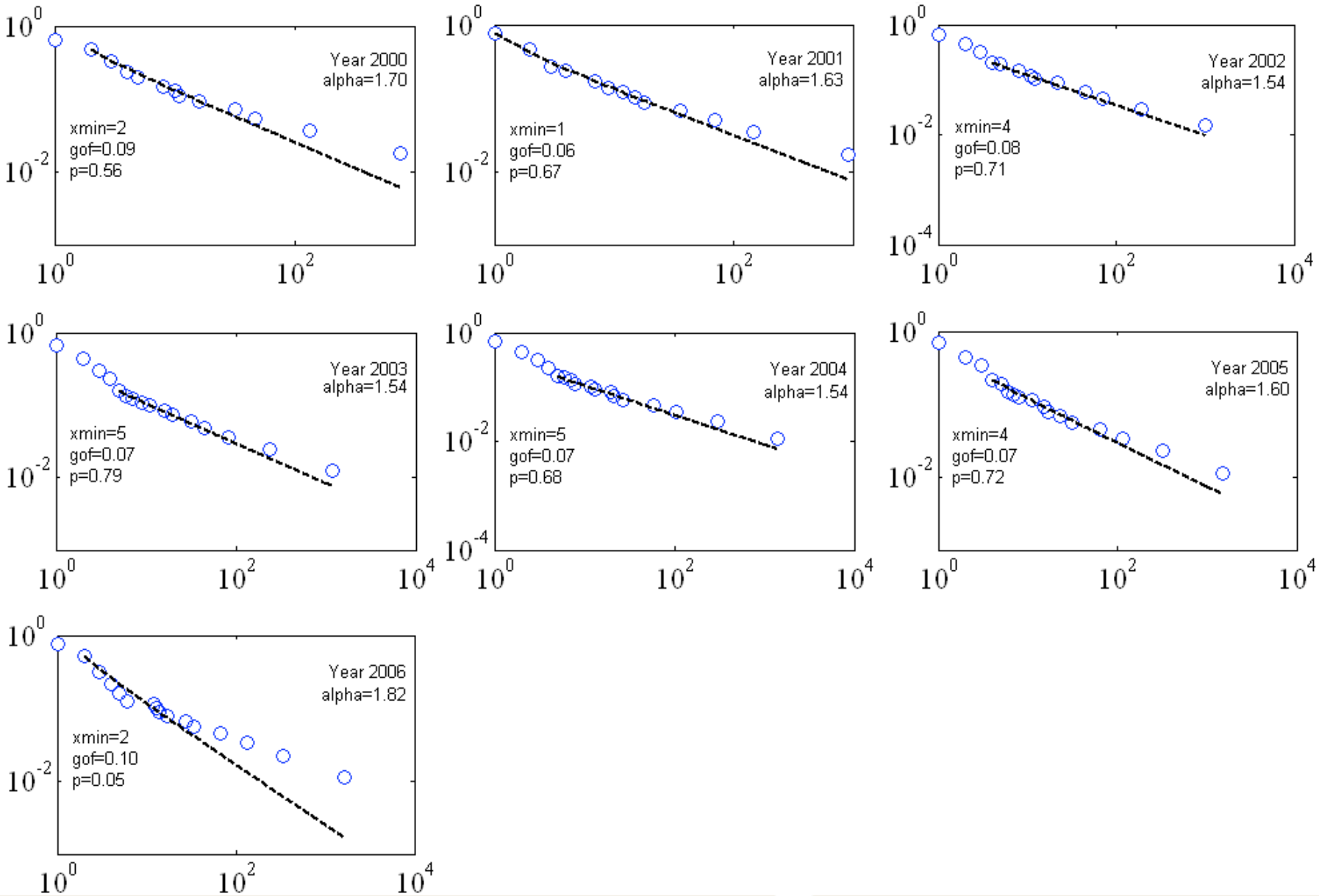
Scale-Free cont.

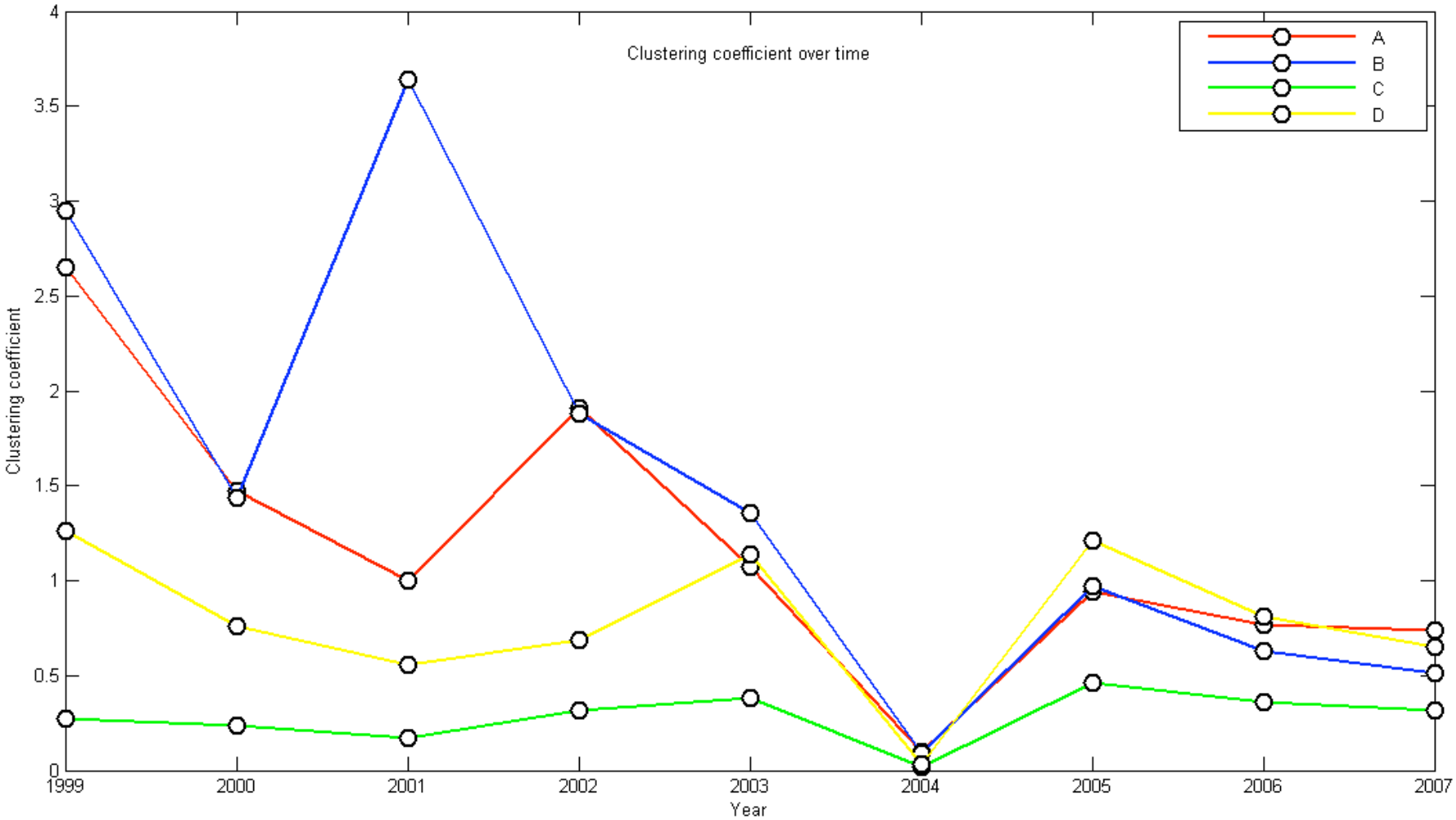
- The diameter of the network (longest path length) should be approximately *log(log(n))* for scale-free networks.
  - **In both cases (for the gangs and the years) the real values are significantly higher than would be expected for a scale-free network.**
- The average path length should be approximately log(n) for scale-free networks.
  - **For both the gangs and years data it was actually smaller than log(n), indicating scale-free networks.**

Scale-Free cont.

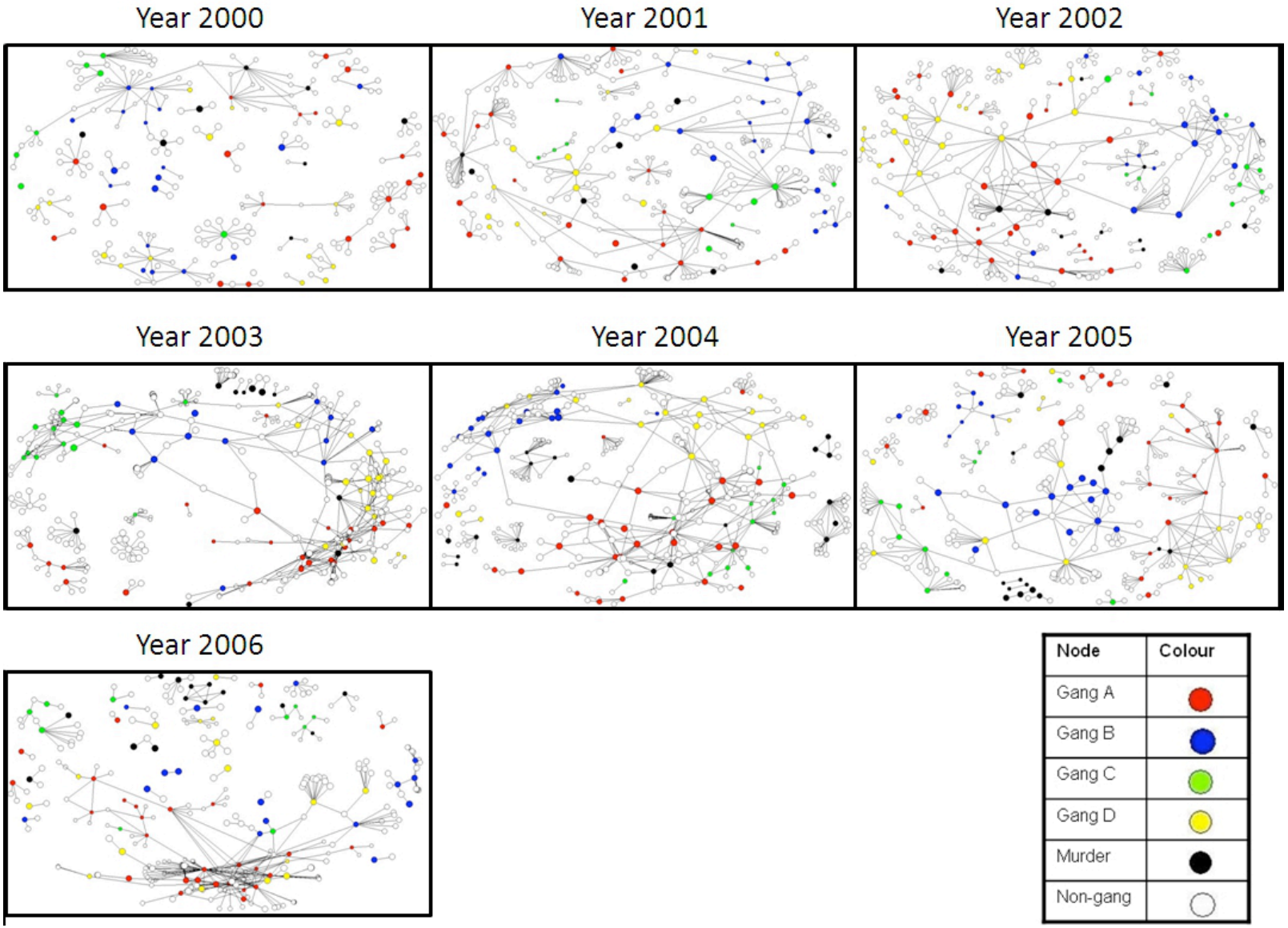
- The diameter of the network (longest path length) should be approximately *log(log(n))* for scale-free networks.
  - **In both cases (for the gangs and the years) the real values are significantly higher than would be expected for a scale-free network.**
- The average path length should be approximately log(n) for scale-free networks.
  - **For both the gangs and years data it was actually smaller than log(n), indicating scale-free networks.**
- The statistics on degree centrality were low, indicating that there is no group leader.
- As we know when Gangs C (2001) and D (2004) are formed, it is interesting to note that the characteristic of the networks at this time are that the betweeness centralisation reaches 0.2.
- It is necessary to compare the closeness and betweenness averages for each gang against the value for the overall network.

Power Law Investigation





Per year clustering coefficients for each gang. Gang C was formed in 2001, Gang D in 2004.



Annual links formation. Only nodes directly connected to a gang member are included.

Conclusions

- Identifying changes can help us identify the possible birth of new gangs (sub-networks) in the social system.
- Through studying the dynamics of these networks globally and locally, we have identified the global characteristics that tell us that they are not random graphs – they are small world graphs – implying that the formation of gangs is not a random event.
- However, we are not yet able to conclude anything significant about scale-free characteristics due to insufficient sample size.

Future Work

- Further pre-processing of original data is being carried out.
- The quality of the data collection process is to be improved.
- Criminal behaviour (modus operandi and offence profiling) is to be incorporated into the social network analysis.
- **There is significant future work available with this dataset.**

Acknowledgements

- Acknowledge contributions from:
  - UK EPSRC Sandpit on Gun Crime
  - Xcalibre Task Force, Greater Manchester Police, UK

- Papers and presentations:
  `https://github.com/tomcrick/ASONAM2014`
  `https://github.com/tomcrick/FOSINT-SI2014`

- Contact: `http://drtomcrick.com`