

Middle of the (by)line: Examining hyperauthorship networks in the Human Genome Project

Ly Dinh, Yi-Yun Cheng

School of Information Sciences, University of Illinois at Urbana-Champaign

I School of
Information Sciences
The iSchool at Illinois

hyperauthorship ; co-authorship networks; author byline

INTRODUCTION

Large-scale scientific endeavors are on the rise:

- e.g. Human Genome Project (HGP), CERN Large Hadron Collider
- Number of authors associated with these research projects substantially increase as well

Co-authorship dynamics has evolved:

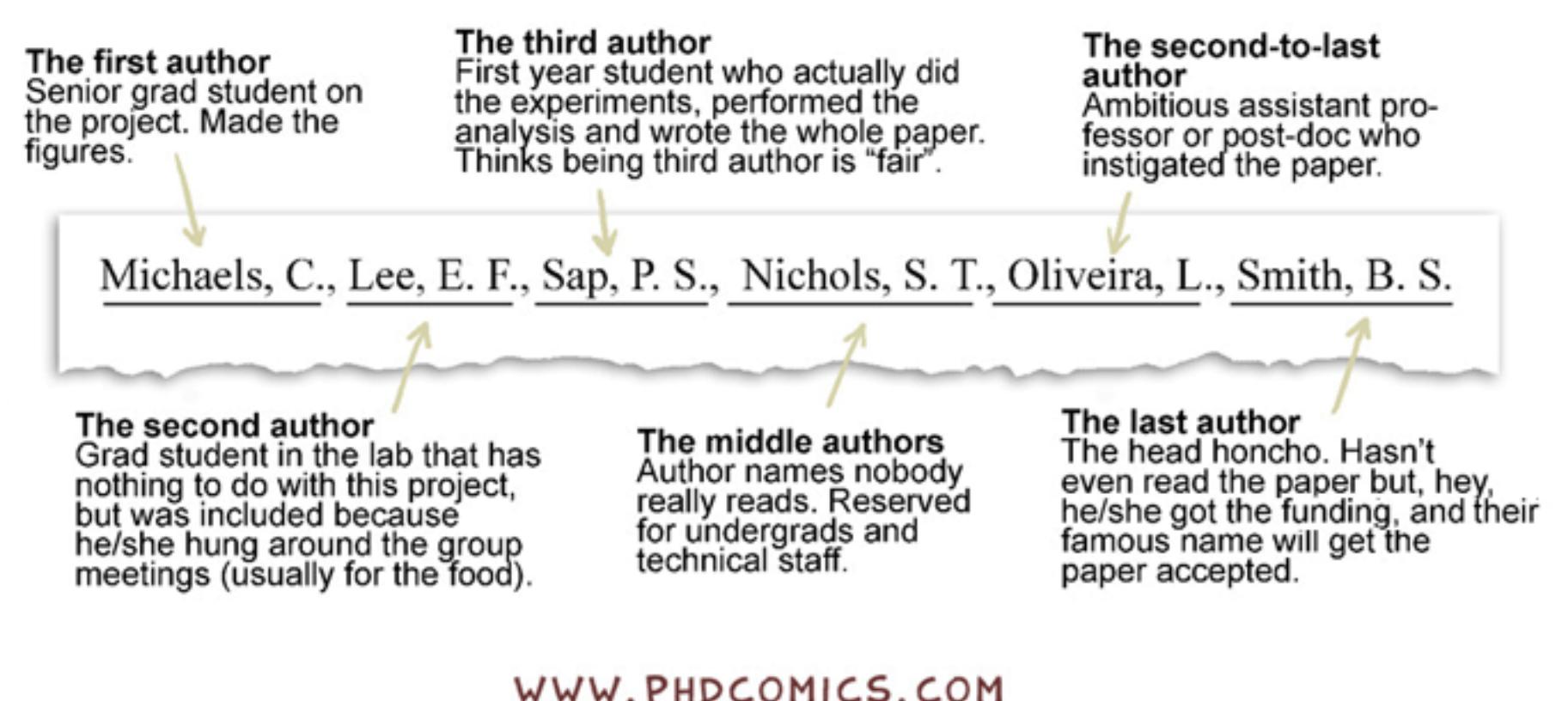
- BEFORE: multiple authors (2~6 authors per paper)
- NOW: 'mega', 'hyper' authors (80~200 authors)

Hyper-authored research:

- More of a convention in scientific domains such as biology, high energy physics, and medicine
- Requires large and complex coordination of tasks

Author-ordering tradition in genomics (e.g. chromosome sequencing tasks):

- Usually listed authors in three parts: first, middle, last
- Group 1 **First authors**: main contributors
- Group 2 **Middle authors**: data collections, annotation → assumed to make smaller contributions
- Group 3 **Last authors**: senior researchers supervising the research



WWW.PHDCOMICS.COM

AIM

RQ: Investigate the **hyperauthorship** phenomenon of one major research publication from the **Wellcome Trust Sanger Institute** (biggest contributors to sequencing human chromosomes)

Using co-authorship network analysis to:

- Examine the collaboration dynamics in HGP
- Extrapolate the three-parts structure of partial author byline (some authors are listed alphabetically, some not)

Table 1. Overview of network characteristics

Measures	Whole network	Induced network
Number of nodes	1918	166
Number of edges	14,088	9,653
Density	0.008	0.705
Degree centralization	14.69	116
Betweenness centrality	2,213	21.63
Eigenvector centrality	0.055	0.557
Clustering	0.843	0.908
Average path length	3.538	1.425

METHOD

The DNA sequence and biological annotation of human chromosome 1

S. G. Gregory^{1,2}, K. F. Barlow¹, K. E. McLay¹, R. Kaul¹, D. Swarbreck¹, A. Dunham¹, C. E. Scott¹, K. L. Howe¹, K. Woodfine¹, C. C. A. Spencer², M. C. Jones¹, C. Gillison¹, S. Sealie¹, Y. Zhou¹, F. Kokocinski¹, L. McDonald¹, R. Evans¹, K. Phillips¹, A. Atkinson¹, R. Cooper¹, C. Jones¹, R. E. Hall¹, T. D. Andrews¹, C. Lloyd¹, R. Ainscough¹, J. P. Almeida¹, K. D. Ambrose¹, F. Anderson¹, R. W. Andrew¹, R. I. S. Ashwell¹, K. Aubin¹, A. K. Babbage¹, C. L. Baguley¹, J. Bailey¹, H. Beasley¹, G. Bethel¹, C. P. Bird¹, S. Bray-Allett¹, J. Y. Brown¹, A. J. Brown¹, D. Buley¹, J. Bygrave¹, C. Callahan¹, S. C. Carr¹, S. Clark¹, R. Clark¹, C. Cleary¹, C. Colley¹, R. Collier¹, N. Corby¹, G. J. Cowen¹, J. Davies¹, M. D. Dunn¹, A. G. Dunn¹, R. E. Erington¹, A. Franklin¹, L. French¹, P. Garner¹, J. Garnett¹, L. Gay¹, M. R. J. Ghazi¹, R. Gibson¹, S. Hammond¹, E. S. I. Harrison¹, E. Hart¹, E. Haugen¹, P. D. Heath¹, S. Holmes¹, K. Holt¹, P. J. Howden¹, A. R. Hunt¹, S. E. Hunt¹, G. Hunter¹, J. Iherwood¹, R. James¹, C. Johnson¹, A. Joy¹, M. Kay¹, J. K. Kershaw¹, M. Kubikawa¹, A. M. Kimberley¹, A. King¹, A. J. Knights¹, H. Lad¹, G. Laird¹, S. Lawlor¹, D. A. Leongamomlert¹, M. Lloyd¹, J. Loveland¹, J. Lovell¹, M. J. Lush¹, R. Lyne¹, S. Martin¹, M. Mashreghi-Mohammadi¹, L. Matthews¹, N. S. W. Matthews¹, S. McLaren¹, S. Milne¹, S. Mistri¹, M. J. F. Moore¹, T. Nickerson¹, C. N. O'Dell¹, K. Oliver¹, A. Palmer¹, A. Parker¹, D. Patel¹, A. V. Pearce¹, A. I. Peck¹, S. Pelan¹, K. Phelps¹, B. J. Phillimore¹, R. Plumbe¹, J. Rajan¹, C. Raymond¹, G. Rouse¹, C. Saenphimmachak¹, H. Sehra¹, E. Sheridan¹, R. Showken¹, S. Sims¹, C. D. Skuce¹, M. Smith¹, C. Steward¹, S. Subramanian¹, N. Sycamore¹, A. Tracey¹, A. Tromans¹, Z. Van Helmond¹, M. Wall¹, J. M. Wallis¹, S. White¹, S. L. Whittle¹, J. E. Wilkinson¹, D. L. Willey¹, H. Williams¹, L. Wilming¹, P. W. Wray¹, Z. Wu¹, A. Coulson¹, M. Vaudin¹, J. E. Sulston¹, R. Durbin¹, T. Hubbard¹, R. Wooster¹, I. Dunham¹, N. P. Carter¹, G. McVean¹, M. T. Ross¹, J. Harrow¹, M. V. Olson¹, S. Beck¹, J. Rogers¹ & D. R. Bentley¹

Figure 1. Three-parts author structure of the chromosome 1 paper

Co-authorship network construction:

- Capture collaboration patterns among all the authors involved in the chromosome 1 study
- Undirected network: co-authorship is mutual between two researchers
- Node i (ego) represents each author; an edge a_{ij} between node i (ego) and j (alter) denotes that these two authors have worked on **at least 5** papers together

Data collection:

- Co-authorship data on articles published by the Sanger institute on the results of sequencing 8 chromosomes: 1, 6, 9, 10, 13, 20, 22, and X
- The 8 papers identified were published between the years of 1999 to 2006, and were all highly-cited within the field
- Focus of this study: published work on **chromosome 1**, the largest human chromosome, thus requiring the most coordinated efforts to annotate and sequence

Unit of analysis:

- Chromosome 1 paper (Gregory et al, 2006)
- 166 authors: **24** are **first authors**, **127** are **middle authors**, and **15** are **last authors** (Figure 1)
- Data on authors' affiliations, general demographics as well as the list of co-authors are obtained from the **ISI Web of Science** and **Scopus**

Tools used:

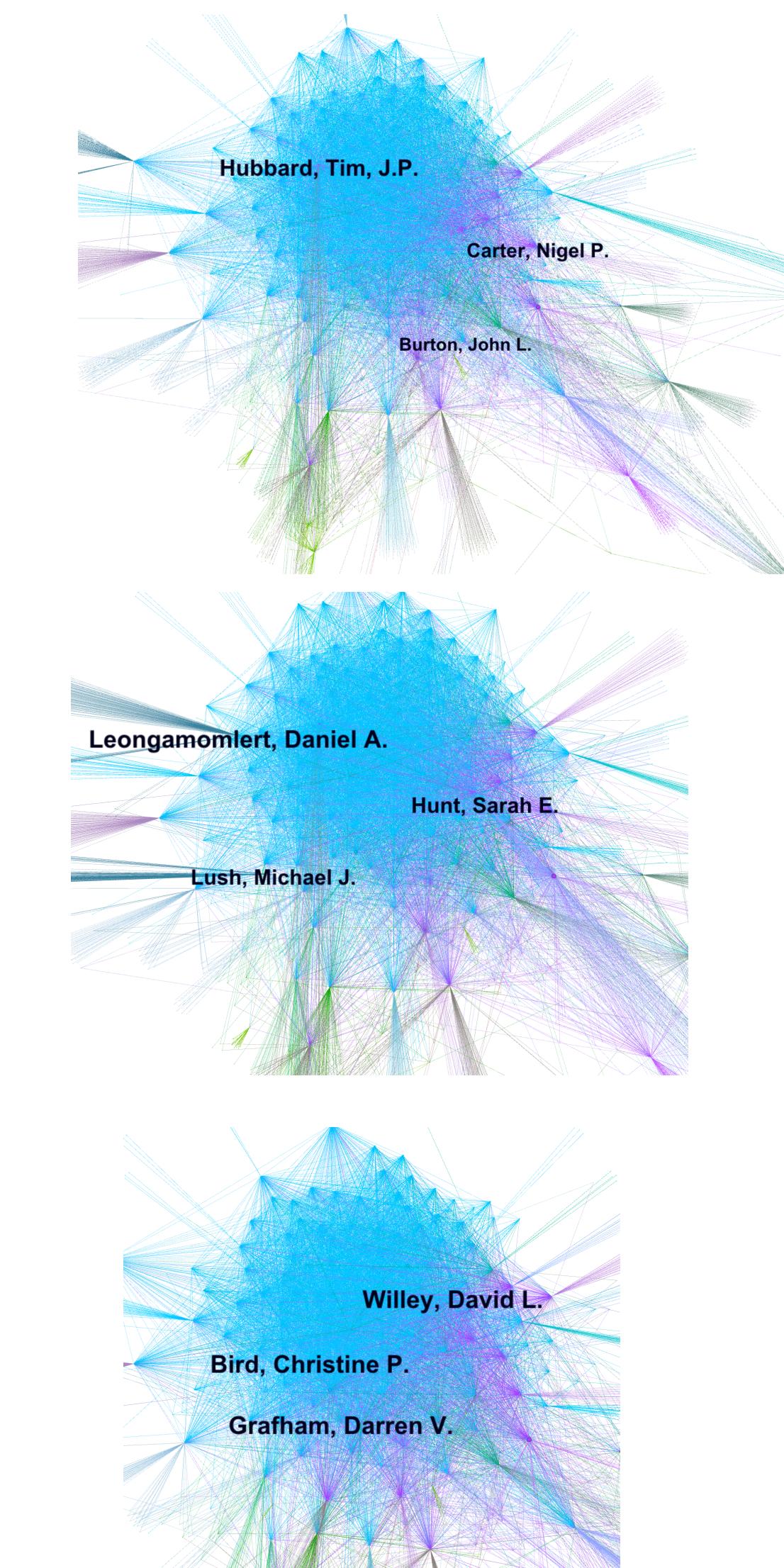
- R's *igraph* package: constructing the network matrix and calculating network measures
- Gephi: visualizing the whole network and the induced subgraphs

RESULTS

Table 1 exhibits the network characteristics of both the whole network and the induced network

The whole network:

- includes ego-alter relationships and alter-alter relationships (co-authors tied to the focal author are also tied to each other)
- Density is low ($p=0.008$) (mean degree of 14.69 edges per author); Clustering coefficient is high at ($C=0.843$); Average path length is low ($\ell = 3.538$)
→ **Middle authors (Group 2)** score high in Betweenness and Eigenvector centralities



Whole- Degree centrality

Name	Group
1.Hubbard, Tim J. P.	3
2.Carter, Nigel P.	3
3.Burton, John L.	2

Whole- Betweenness centrality

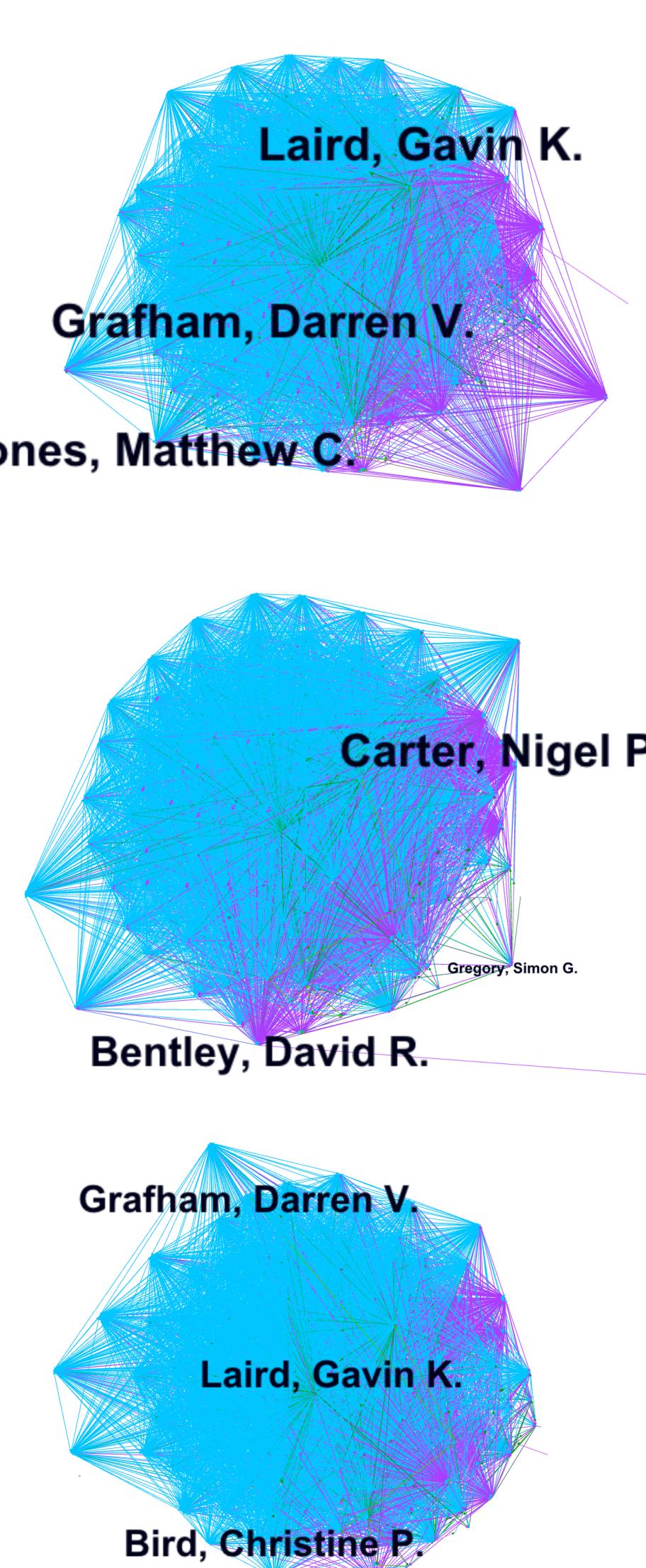
Name	Group
1.Leongamomlert, Daniel A.	2
2.Hunt, Sarah E.	2
3.Lush, Michael J.	2

Whole- Eigenvector centrality

Name	Group
1.Graham, Darren V.	2
2.Willey, David L.	2
3.Bird, Christine P.	2

The induced network:

- subset of the whole network with specified egos and alters who are 166 authors in the chromosome 1 paper
- Density is much higher ($p=.705$) (mean degree of 116 edges per author); Clustering coefficient is very high at ($C=0.908$); Average path length is minimal ($\ell = 1.425$)
→ **Middle authors (Group 2)** score high in Degree and Eigenvector centralities



Induced- Degree centrality

Name	Group
1.Graham, Darren V.	2
2.Laird, Gavin K.	2
3.Jones, Matthew C.	1

Induced- Betweenness centrality

Name	Group
1.Bentley, David R.	3
2.Carter, Nigel P.	3
3.Gregory, Simon G.	1

Induced- Eigenvector centrality

Name	Group
1.Graham, Darren V.	2
2.Laird, Gavin K.	2
3.Bird, Christine P.	2

CONCLUSION

- HGP is a leading example of big science that requires collaborative efforts from hundreds to thousands of researchers who are from a wide range of disciplines

CREDIT WHERE CREDIT IS DUE

- Determining where credit is due remains a conundrum in a hyper-authored setting
- Middle authors** hold essential positions within the collaboration network in the chromosome 1 research
- Misallocation of credits may result in misallocation of funding opportunities and academic positions

→ **OUR VIEW:** Providing a **contribution list** to describe **whodunwhat** in conjunction with the author byline is a necessary practice

Future work:

- Examining in detail the collaboration dynamics among the first, middle, and last authors of ALL published works by a number of research centers affiliated with the HGP
- A post-hoc analysis reveals that the topology of this network exhibits more desired properties of a **small world network**, rather than a scale-free network as with most co-authorship networks (Barabasi & Albert, 1999). We will also examine this finding in further details.

REFERENCES

- Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the Association for Information Science and Technology*, 57(13), 1758-1770.
- Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices?. *Journal of the Association for Information Science and Technology*, 52(7), 558-569.
- Gregory, S. G