Transcript of a virtual meeting held via Zoom on Monday August 8, 2022. The purpose of the session was to discuss the results of this dissertation in the context of historical ground-truth information contributed by the Co-Investigators on the ToK project.

Participants:

Nayomi Kasthuriarachchi (NK): Dissertation author

Dr Duncan Hay (DH): Dissertation supervisor. Also Research Fellow/Developer on the ToK project and Research Associate at CASA

Dr Alex Butterworth (AB): Co-Investigator on the ToK project and Research Fellow in Digital Publishing, Data Visualisation and History at the University of Sussex

Dr Rebekah Higgitt (RH): Co-Investigator on the ToK project and Principal Curator of Science at National Museums Scotland

Dr Boris Jardine (BJ): Co-Investigator on the ToK project and Researcher at the University of Cambridge, working in the Department of History and Philosophy of Science and the Whipple Museum of the History of Science

---

DH: [0:03] Thank you, Nayomi. That's great. Okay. Should we kick off? And then when Boris turns up we could... Oh, I need to let screen sharing.

NK: [0:17] Yes, please. Thank you.

DH: [0:19] And here is Boris.

NK: [0:22] Perfect

DH: [0:32] Hello, Boris.

BJ: [0:33] Hi. Sorry about that. Little bit hot, having cycled across town.

AB: [0:40] You look hot! Thanks for kind of hastening on here. So, yeah, we're just doing introductions. So if you could pant out something about yourself.

BJ: [0:48] Yeah. Hi, I'm Boris. I am a Co-I on the Tools of Knowledge project. I don't know what you need to know. I've been working at the Whipple Museum for a long time, working on various aspects of the scientific instrument trade, which is what led me to the project. And I don't think you need anything else, necessarily.

NK: [1:10] That's fine. Thank you very much. I am Duncan's Master's student over the summer. And I've been looking at the Tools of Knowledge dataset. I'm just fascinated. I'm looking at it from an algorithmic community detection perspective. So what I've done is just I've run one particular type of community detection algorithm on the dataset that maximizes a particular function called modularity. And it's coming back and telling me that there are over 1,000 communities as it has determined within the data. So I'm interested in knowing... Well, the algorithm tells me that these are communities. But what does that actually mean? And by virtue of running this process

through a graph database, my aim is to try to make my analysis richer than perhaps other community detection projects, by being able to say not only we have communities and how many there are and some very high level things about them, but really be able to look within these communities and look for patterns of participants, guild memberships, locations, basically whatever I've been given through the data, I'm able to pull in. So yeah, just trying to make sense of what I've got. For today, I've run 1, 2, 3, 4. . . 4 broad categories of analysis. So first of all, I looked at the five largest communities. Secondly, I looked at the evolution of a particular community. So how that grew through time. I then looked at... say that we were interested in a particular product. So in my case, I've just picked out umbrellas just because it was one of the smaller products in the database, it just seemed more manageable. If we start with umbrellas as a product, looking at which communities made that product and how they might be shaped and what I could see from that. And then I also looked at, I picked up particularly influential individuals, and the way that I defined influence in my work was using the Google PageRank algorithm to rank individuals according to the algorithm. So I'm just trying to decide where to start and if any of this is going to be useful at all, I could start with...

AB: [3:34] Could I just ask? Did you say something about the Google. . . ?

NK: [3:36] The PageRank? So I just use that just as a method of looking at which of the nodes in the graph it. . . and it's not the only method but it's the method that I used to determine influence within the graph. So I'm really trying to decide where to start. We can start looking at the five largest communities and the shape of them because there were some interesting shapes in there. I'll try to share my screen, so bear with me while I try to work out how best to do this. So, ignore that for a second. . .

AB: [4:19] Does Nayomi have screen sharing allowed?

NK: [4:21] Yes, I can you not see? I'm sorry. I thought I was sharing. It says that I am.

DH: [4:35] That's weird.

NK: [4:36] Oh, it wants me to pick. Hang on. Let me try and share. Now I can't see what I'm sharing. Hang on.

AB: [4:47] You use your email from Duncan. So yeah.

NK: [4:50] What can you see now?

AB: [4:52] The Neo4j graph with a dialogue to open a [inaudible]

NK: [5:02] Okay, so I'll start with the biggest one, then. Sorry, it's going to be a little bit awkward. I couldn't figure out a good way of doing this. So this is the largest community. So it's got 172 members in 25 locations, 75 guilds and 16 products. Just trying to learn how to make this... how to navigate around this easily. It was a very London-centric graph. I mean, some influences from as far as St. Petersburg, apparently, through this J. Parminter. So I should just explain here the orange bubbles are individuals, I've put in first initial and surname. And the number that you see there is the identifier from the

database. I had to do that because there are individuals who are named the same, it got really confusing. And I just needed to make sure that it was talking about different people. The beige-y circles are the products made, and the pink circles denote guild memberships, as far as I'm aware. I don't really know... So just tell me if me moving around like this is helpful. It's... I can't really figure out another way to give you a view.

BJ: [6:24] So wait, what are we looking at here? We're looking at one of the largest communities?

NK: [6:29] We're looking at the largest community as defined by the algorithm.

BJ: [6:34] I mean... I know that you can't give like an introduction to the whole black box stuff. But is there any way of explaining what you think it has done in order to select this group?

NK: [6:47] So it goes through the nodes and it looks at how many connections each node has and then it pairs them depending on looking at the neighboring nodes and saying, right, "You're in a community with me." So that's the first pass, then it will go through the second pass. And what it does in the second pass is the communities that it put together in the first pass, it makes them larger. So it put two individuals together in the first pass, they then become one community for the second pass. And it keeps on going through like that, until it can no longer maximize the underlying function that it's trying to maximize. And then...

BJ: [7:24] Would I be right in saying that the only node type that it is counting... are individual people?

NK: [7:30] Yes, I should mention that. It's only running on individuals. The guilds and the locations, I brought all of that in afterwards. So this is all based on relationships that individuals have, between each other, in terms of determining a community.

BJ: [7:46] And are the placing of the nodes or the points here? Is that significant?

NK: [7:53] I think it is. I'm not clear, because this is generated by Neo4j using its... like its force directed layout. So I think what it pulls together in the centre are the ones with the most connections. And then on the periphery, you get the ones with the fewer, the fewer connections.

BJ: [8:13] And when you're running the algorithm, the algorithm that does the visualisation, is that including places and guilds as nodes?

NK: [8:24] I'll explain. So what I do is I create a graph in memory. And I say, "Run community detection on this graph", which is just all the people in the graph. So then it creates a graph for me in memory. And when I want to do a visualisation of a particular thing, what I say is, "Okay, give me back all the people. And where you can, give me back where the people were known to be active in, in terms of places, and where you can give me back the guilds of which they were members". I can make any configuration. So this isn't the only one. The only thing I need to do in order to generate the communities is to have a graph in memory of the relationships between people. Everything

else is optional. So I could have shown you a graph with none of the locations and none of the... the guild memberships, for example, and then you would just see the people who are connected.

DH: [9:17] Just looking at this. So I'm seeing a beige node with 'globe' written on it. Is that highly connected in the...?

NK: [9:28] Yeah, so they made globes, and it's central to the graph. And my thinking is that all of those people around it, I could go in further to investigate it by calling a separate query. But my thinking is that it's so central to the graph because it has so many... it's connected to so many individuals and hence the reason it was placed there. Compared with something like dial out here because it was only made by this individual.

RH: [9:55] I saw that... feels a bit odd because I saw 'rule' at the bottom and you would think there... will be an awful lot of makers who made rules compared to globes?

BJ: [10:03] Yeah, did Joseph Moxon not make globes? I mean dials. See what it says in the database.

AB: [10:08] Joseph Moxon is close to 'globe'. Senex and Moxon are both quite close to 'globe'. I don't know whether that's sort of indicative of...?

NK: [10:27] So, did this person make rules as well? Because it doesn't seem to suggest from here...

AB: [10:34] Can I ask... At a glance, do you see anything about these... I mean, my impression is that the dates of these makers are quite broadly spread.

NK: [10:45] They are. So this is not stratified by date. I do have the...in terms of date, I've been able to extract the first date of activity, commencement of all the individuals, I have the first date, and then I also have the last date so I can give you a span. What I've also been able to do when we go to the community evolution is I've been able to... I can split up the time span, for example. I can take every decade and look at how the community came together, which is something we can...

AB: [11:22] That could be very interesting. Anyway what I'm thinking here is that although the dates are not kind of active within the graph, you would expect that if... Well, I mean, in a sense date to some extent, likely to be a proxy for personal association. So since date is not so prominent here, it may be the dissociation is rather kind of thematic either through association with kind of persistent guilds, or instruments made, which will also have a persistence, and I'm just... that's just pure speculation.

NK: [12:07] I mean, one of the frustrating things here is I really wanted to be able to differentiate person nodes by dates, but I haven't... it's not easy to do and I don't even know if it's possible to do it. And I haven't been able to do it at this stage, but it's something that I would like to do. And then that would give us an idea of... time in which these individuals were active.

AB: [12:28] And Neo4j, it doesn't allow any sort of optimised layouts beyond this, does it? I mean, it's basically... ego network.

NK: [12:39] Yeah, this is what it gives me out of the box. So if we can look at... try to do a different one... there was one with a very distinct... So this one I thought was quite interesting. This is the third largest community: 129 members and 26 locations. So here, there was quite a distinct Scottish community down here. Who... they seem to be associated with compass adjusters. Yeah, so it's a Glasgow-centric community. And also Liverpool down here as a cluster, which ultimately... and then up here is London, but it's this one doesn't seem to be as London-centric as the first one.

AB: [13:46] That's quite interesting. Boris, from your perspective of and Becky's with the Scottish angle but also the kind of industrial centres or provincial centres of activity.

BJ: [14:01] Yeah, it's interesting that it's... I mean, it's actually North and South Shields have slightly kind of jumped apart from each other. But also these are later makers, I think. These are 19th century makers. I'm pretty sure.

NK: [14:19] Yeah because it's not differentiating by time... it's just every individual in the database.

BJ: [14:23] Yeah. But I think it would only be possible to get a kind of multi-city network later in the data because you just wouldn't have connections. I mean, it's... it'd be interesting to see how. So I have to keep reminding myself that compass isn't actually acting as a node that connects Glasgow and London.

NK: [14:41] No, it's not.

BJ: [14:43] So you can see the ones that make the jump like Gardner, Kelvin. Yeah, some of the others.

DH: [14:55] It's all really interesting isn't it? Just looking... so who are they...so there's a distinct Glasgow cluster and a distinct London cluster. It seems to be like a small number of makers that are bridges those two communities. But so is it Gray.... No, J Sewell is sort of in the middle next to the spectacle makers guild.

BJ: [15:20] Yeah, I think with this one, what you're seeing. My guess with this one is that what you're seeing is actually a much smaller and much more tightly connected group of instrument makers because if you notice all of... anything with either name Kelvin or White in it is actually one... that's one concern because White was the instrument maker for Kelvin. And W. Thompson is Kelvin. So there's kind of duplication, which is not a bad thing, it just means that that's actually like one... and you can see loads of Gardeners. So clearly, that's a family of Gardeners. So probably, there's like, the migration or movement or collaboration of one or two families or groups.

AB: [16:11] And it is the... one of the interesting challenges, I think that this is to... Well, it's one of the things we're wrestling with is...the different ways to identify the identity of Maker. So is it an individual? Is it a going concern that stays within the family and does sort of relatively consistent things? If you view it in those terms, do you get a very different picture?

BJ: [16:38] Yeah, it's tricky if you're thinking about significance because obviously

the families are kind of going to win aren't they because they're multiplying... the same entity and also any business that has multiple stages, like & Co and & Sons and so on, will probably retain the same connection. So those connections will look twice or three times as important.

RH: [17:00] Yeah, I mean, this one's very much about business, succession and name changes and that sort of thing whereas the other one..that much more about apprenticeships, which, therefore connect with livery companies, which is going to be London.

BJ: [17:16] Yeah, that's true.

DH: [17:18] That's really interesting. So you, Boris and Becky, you can you... can you begin to see, you can see things that are characterising these communities that make a coherent sense to you?

BJ: [17:38] Yeah, so in the terms that I just said that it's... and that Becky said in a different way that I think what we're able to see is why it's grouping these people. But that's only because of our kind of blinkers that we just... you'd notice that of the oddities of the data that would produce something like this, but that... for example, I don't know who J Sym is, or why he would be there. So there's a lot of things that I'm obviously missing as well.

RH: [18:06] But there's J Sym and J Sym & Co and J Sym, different version.

BJ: [18:11] It's all under there. Yeah. But this is what I found, when I did the stuff in Gephi that you could get quite fine. You generate these really interesting patterns. And at a certain point, you would start to understand why they were developing as artefacts of the way the data is either recorded or where it comes from, which is... so again, it's no bad thing is, it means that, at least this first pass that I did, I was kind of felt like I was analysing the database rather than analysing the instrument trade, which is two slightly different things.

AB: [18:51] So how do you get beyond that, Boris? What's the next step that's needed to...? What obscures the seeing beyond the structuring principles and the selected principles?

RH: [19:08] I think bringing chronology in is always going to be helpful. Slice it all... or whatever.

BJ: [19:17] That's definitely the one I think. Because then you can do comparisons. So you can say, well, I don't care that it's throwing up some artefacts, but I'll look at 1700 and 1750 and see kind of what's different across those two. For however, I was like, faceting.

NK: [19:38] Okay, if we go to then...

RH: [19:42] Sorry, one thing before we go... I'm seeing something we discussed before is quite a lot of 'See Also', which is often an artefact of actually knowing very little as in, this person has the same surname as someone else. They could well be related, we really don't know. And it may be no more than that.

AB: [20:00] Yeah, stripping out the 'See Also's is probably quite a good thing to do.

BJ: [20:06] When I got into the...the further I got, the further I wanted to segment the relationship type. So what I would do is color them differently. And what I would do is not color them just across the whole palette but I would categorize... I'd clump them. So I'd have like apprenticeship type relations or employment relationships, I'd make them all blue, and then I'd make all family relationships green or something. And then all random relationships, I'd make grey. So I almost couldn't see them because they were just not so helpful.

NK: [20:36] I wonder if that is possible. I know I have segmented the family relationships from the others but not in here. I'm just wondering if it'd be possible for me to do that in here. I'd need to investigate that.

AB: [20:50] How easy is it, Nayomi, to export these filtered graphs into a form that can be analysed in Gephi, for example?

NK: [21:00] I think there is a bridge, isn't there, Duncan? I think between Neo4j and Gephi?

DH: [21:05] Yeah, I think you can definitely make them talk to each other. Yeah, I think as long as... I think that would be very possible.

NK: [21:14] Yeah. I'd be really surprised if they hadn't made that possible because it must be something that people who work with these graphs all the time need to do.

AB: [21:23] I think that will be... I mean, if you have time, I think it would be well worth doing that. Because I think you would get a much more kind of malleable, optimised view of the data. Because... there's a lot that needs to be unpacked, which is essentially what [inaudible] are doing.

RH: [21:45] The one other thing to say about this, I think we've got quite a practical, professional kind of cut. I mean, there's quite a lot of shipping support kind of objects here. And that would be to Glasgow and Liverpool. That's what we would expect.

DH: [22:06] That's really interesting.

RH: [22:07] That looked different to the previous one, I certainly think.

NK: [22:15] So if I look at...

AB: [22:22] We're also talking about [inaudible]...I'd be very interested even as someone without the domain expertise to kind of get it and just start teasing it apart instead of looking at it or being able to look at this filtered data subset in different ways. Which would sort of facilitate giving, perhaps more useful feedback.

NK: [22:43] I'm just trying to see if I can show you. So in terms of... just looking for it. I can actually flick through.

RH: [22:57] Those ones that seem to loop back on themselves, I saw that as well.

NK: [23:04] I have found that in the data. So there's somebody that was 'child of' and it loops back on themselves.

BJ: [23:12] There's one case of someone taking their father as an apprentice. Might be a watchmaker.

RH: [23:21] That would suggest that maybe we haven't done our disambiguation fully.

BJ: [23:26] I don't think you can take yourself as an apprentice unless you're doing something like sort of setting up a front for drug running.

DH: [23:33] Sounds a bit like money laundering.

NK: [23:37] So this is my visualisation. I picked a particular community that was small enough to break up like this. So it's a community that the first entrant was recorded and... I think the first activity was 1763. So, I tried to break this down by decade and look at how the network then evolved. So, this is 1760-1770. And there were two individuals active here. They're the orange nodes here and I brought in location and guild.

AB: [24:17] And so, this was a small community for this time period generated with those attributes, with that limited [inaudible].

NK: [24:26] Yeah, so, I picked a community that was small enough for me to then be able to expand and look at the changes. So this is up to 1770, then if we move through this is to 1780, we have two more entrants and then I can keep on... I kept on expanding it. So this is up to 1810 now and this disconnected part will eventually join the network. That was the next... so I just split it up and gradually grew it. I don't know if this is an interesting way to cut the data or not.

BJ: [25:08] Very. This is really good.

NK: [25:11] And so we just keep going through like that. Rolling it. I mean, the one thing I would like to be able to do, again, is differentiate the nodes. So we know to which period they belong. But I haven't been able to do that.

AB: [25:24] Yeah, well, we're looking at in terms of bespoke visualisation, we're looking at something that will present this kind of evolution. Something more like kind of a business genealogy, using space for chronology but also kind of coloring by different attributes.

NK: [25:43] So yeah, so this is the result. So when there are no more entrants... so this was 1890. The community as determined by the algorithm is complete.

BJ: [25:56] All members have gone from that. Well, the original members must have gone there.

NK: [26:02] Well, that's the thing. I don't remove them. Because if I remove them, it breaks the links. So I've had to preserve them, which is why I would like to ideally be able to show time period. Because I think one of the downsides, it leads to the false impression that these were all active concerns in 1890, which they weren't. But I did try to remove them. And it just did...it broke up the network.

BJ: [26:28] You can almost imagine it flowing through time, imagine the way it would look if it...

AB: [26:33] I mean, purely as kind of contribution to the design process. Having this and a handful of similar examples would be really useful. Having both as visualisations but also as exported graphs, would be very useful in helping us think through.

NK: [27:00] So yeah, I mean, this was the best I could do in Neo4j for the moment. But I found it useful, just to give me an idea of the changes over time, and how the ultimate network, how it ultimately evolved.

BJ: [27:14] Quite a lot. I mean, you can see, for example, that the guilds that are involved are established early, and probably after, like, 1800, or whenever it was, they don't. You can add more people without them having to be in a guild. So that's self-evident.

AB: [27:30] And what about the... Can we go through and see just that through again?

NK: [27:35] So this is 1770, 1780. Two more entrants. By 1810, there were seven. And the reason I haven't gone... because in some decades, there were no changes. So the next change happened here. 1820, there were eleven.

AB: [27:59] It's interesting that... what happens to the instrument types in the previous one?

NK: [28:05] Because you don't always have an instrument type associated with a person. So it's where they are available. I think that's right. Yes. Because Freeman is only here. Now at 1820 and Freeman had a record in the database of making an instrument and where they are available, I bring it in.

AB: [28:28] The instrument times drop off as the maker actually... The maker stays in the graph.

NK: [28:32] No, if the maker had an instrument type then I will bring it in. If this is not there, then I can't.

RH: [28:42] So there are no instrument types in this graph until this point, it's not there at the beginning. They haven't dropped off.

AB: [28:49] Sorry.

BJ: [28:50] I thought it's a little bit odd because W Cowland. I'm looking at the right person. He does have an instrument. He's got a quadrant.

NK: [29:01] So what I found in the data or the particular sheet I was using. There were two columns so there's maker descriptions and makers of products. So I think this is just one column that I've brought in. It could be information is in another column, I should really bring them both in. There was a column called Maker Descriptions and a column called was it... Maker Of? But it seems that you could have a Maker Description and not have a Maker Of.

BJ: [29:35] It might correspond to advertised instruments and known instruments because Cowland doesn't have any advertised instruments. He's just got a quadrant.

DH: [29:42] I think Nayomi's been working with cleaned advertised trades.

AB: [29:49] So that's why we got the strange turning of the [inaudible], but yeah, for the [inaudible], I suppose. Yeah.

BJ: [29:57] I mean, it's amazing that he ends up being at all well connected, I suppose. Actually, no, it's true. He had loads of apprentices.

NK: [30:05] Oh. Yeah, we see one, two, and then one to himself.

BJ: [30:10] Yeah, that is weird... Oh, no, it's not him. It's his son. His son is also called William Cowland.

AB: [30:18] You shouldn't be linking to the same ID, then.

BJ: [30:20] No, something's gone wrong, but that's why it's gone wrong.

RH: [30:24] That's one to correct.

AB: [30:28] Can we just step through a couple more now?

NK: [30:30] Yep. So this is 1830, 12 active members. Well, 12 members of the community, they may not necessarily all have been active at this point.

DH: [30:43] So what do you think the uniting factor might be here, Boris and Becky?

BJ: [30:53] It's very hard to say.

AB: [30:54] I might suggest that..the reason I suggested stepping through this is even if only some of them have advertised trades... if they aren't changing. No, they aren't changing now... Ignore me. Because they're not coming in. Are any more instruments introduced in the course of the...?

NK: [31:11] We can take it... So we've got four here. Yeah, here syringe and barometer. [Inaudible] same maker.

BJ: [31:22] If I were to hazard a guess it would probably be Cowland and his 10 apprentices. Because if you have 10 apprentices, then you're going to...

RH: [31:32] But interestingly, it's only [inaudible] two and a half of those.

BJ: [31:37] Yeah, right. So maybe a group of people with an unusually large number of apprentices, but I haven't looked at the others.

RH: [31:49] But is it likely that this would exclude people who are connected to someone in a group like this? Would it not bring in everyone who would seem to be connected?

NK: [32:00] That is what we're doing.

RH: [32:02] Yeah. But it seems to be missing off several apprentices that he did have that we can't...

BJ: [32:08] Doesn't it draw a boundary and it doesn't include people who are themselves not contributing to the "community-ness" of the community?

NK: [32:18] I'm wondering if I've got... if I actually have all of the apprentices in my data.

AB: [32:23] I think you would have. Yeah, you should have... I think you had all the relations. So you should have all the apprentices. But it'd be perfectly reasonable for an apprentice not to have the level of connected... of cumulative "connectedness".

RH: [32:38] They only have one connection, therefore they dropped off?

NK: [32:41] Possibly they then went on after... I don't know if this is... this is pure speculation... I've got no idea. But perhaps they then became more connected to other individuals and went on to be part of a separate community?

RH: [32:55] Although that said. Well, as you know, maybe it changes over time. Johnson above Cowland at the moment is connected just to him and London. But maybe he gains another apprentice or something later or...

NK: [33:12] We can have a look. The annoying thing about this is it does move the nodes around.

RH: [33:21] He's got someone else now. Yeah.

BJ: [33:27] The problem that I have is I don't know who any of these people are. So I can't go by recognition of anything.

RH: [33:34] I mean, again, I think it's certainly apprentice... Most of the ones I can see look like had "apprentice" or "apprentice to". So it just happens... [inaudible] to be in the centre.

AB: [33:50] Nayomi, I think we spoke about this last time we met but the communities detected can they be... are they absolutely mutually exclusive? Or can they be overlapping or subsets of one another?

NK: [34:02] Oh, yeah. So I think there is... I don't think this detects overlapping communities. Yeah, I would have to check, which is an issue with this kind of, I think.

AB: [34:17] Okay. Yeah. Okay. So they have to be discrete?

NK: [34:25] But I do believe that there are methods for detecting overlapping communities.

AB: [34:33] Pretty interesting. Tantalising. I mean, it's interesting, but it's tantalising. Because you want to be able to kind of tease apart and get and see different facets of the information around these to be able to interpret..I think.

DH: [34:50] Nayomi, could do... could you talk us through some of the PageRank results?

NK: [34:54] Yeah, so I'll put it up. I think I only took a screenshot of the top 10. But I can run however many if I open... Yeah, so this is what it gave me as the top 10 most influential individuals in the graph.

DH: [35:33] Well, there's Jesse.

BJ: [35:34] There's some familiar names there.

DH: [35:38] It's interesting that Jesse's not in the biggest community.

NK: [35:44] Yeah.

DH: [35:46] Sorry, are they ranked by the size of the community?

NK: [35:49] No, sorry. This is ranked in... by PageRank Score. So by how influential they were, as determined by PageRank.

AB: [36:00] What does PageRank do in this sense, this has to do with the number of outward connections, is that right?

NK: [36:08] It's looking at the number of connections that a node had and also the number of connections that a node's neighbour's had. So if a node is connected to other influential nodes, it itself then becomes more influential. It's not purely determined on how many nodes are coming in and out. It's also looking at how many are connected to that particular node.

AB: [36:27] To one degree or more or potentially multiple degrees?

NK: [36:30] I believe... it is to multiple degrees.

DH: [36:35] That's interesting, because that's given us a different result to just degree.

NK: [36:42] Yes. Degree looks different.

DH: [36:46] Because if you do it just by degree, Jesse Ramsden's right at the top.

NK: [36:52] So I suppose which is the more useful measure? I couldn't decide because there are so many ways of looking at...

BJ: [37:02] I mean, they both have thrown up the same kinds of people, I think I'm right in saying... these are all familiar names. Actually, maybe not after Cook and Wellington. I don't think we have seen Casella although obviously that's a famous name.

AB: [37:18] Who are Wellington and Cook?

BJ: [37:19] I don't know. I think I've seen all the others at least. Stedman... Nicholl... I don't who Nicholl is. White and Harris are slightly kind of artefactual ones because they are... They're very well connected, owing to bankruptcy proceedings: they both went bankrupt.

AB: [37:45] So the historical record shows them having more connections because they're...

BJ: [37:49] Which is itself kind of amazing that just the fact just that like... lifted the stone and there's all these worms crawling around for those two. Would be the same for anyone if you just knew, but you don't know, which is a bit frustrating. Bate and Ramsden are not surprising because they're both known, well connected and have been largely studied.

AB: [38:14] Are they roughly contemporary?

BJ: [38:20] Roughly. [Inaudible] is the one that just always astonishes me. I've only know him from these rankings.

AB: [38:26] And his date is... Is he early?

BJ: [38:29] Early! He's 1747 to 1774.

NK: [38:34] Sorry, who's that?

BJ: [38:36] Stedman. The thing that happened with him is that somehow someone found out that he employed 15 men in 1759. I don't even know what their sources are for that. Because there's a bunch listed. He's only known to have... the only surviving instrument are coin scales but he did make a full range of mathematical instruments. There's a big family of "Sted-men" and then there's just this colossal list of apprentices and employees. Like it's really, really long.

RH: [39:07] Which company was he?

BJ: [39:08] Stationers [inaudible] in 1745. Educated at Christ's Hospital School as well, just kind of slightly unusual. And then loads of manuscript sources, which I guess is where this all comes from. Well, they're long.

AB: [39:32] So we should do some more digging into Stedman, I think.

BJ: [39:34] Yeah, and I mean, there's secondary sources. Maybe one of them is the definitive biography of Stedman but I don't think so. I don't think this person has ever been really written about which is kind of amazing. So I kind of can't get my head around the fact that this person scores so much more highly than everyone else.

AB: [39:58] Be interesting to look at the... well, when the 15 employees, I suppose as... it must have been Stedmen when I was doing the genealogy that had that kind of vast sort of fan of employees, which was unique at the time.

BJ: [40:16] Again, most of them not in the trade, so they can't be accounting for this. It's not just the size of your workshop. So it's something about... Who are these other people, James Chapman, William Hamlin? I'm just not seeing any other names I recognise here of the people who are also instrument makers... John Atkinson, maybe.

AB: [40:41] I'll spend some time having a look at that. I'm intrigued by that.

DH: [40:48] Which community is Stedman in?

NK: [40:50] Yeah, I was just about to try to call up the graph. One second.

BJ: [40:47] He does have one... His one weird trick is that he was... he sold instruments to James Watt. Now that might be the one that did it.

AB: [41:15] Why would that be?

BJ: [41:17] Because James Watt is mind bendingly well-connected.

AB: [41:21] He sold instruments so he would inherit that kind of network.

BJ: [41:25] Yeah. So Watt is kind of maybe a quite unusual node because Watt doesn't have a very influential workshop in terms of like, being part of the trade because famously he went off to do other things. But he's also because of that very well connected. So he's like a bridge. And he couldn't come up.

Maybe he could, but it seems... doesn't surprise me that he's not there. But maybe someone associated with him is, I don't know.

AB: [41:48] I'd be fascinated to know whether there are other people akin to Watt in those terms, who Stedman is associated with.

BJ: [41:59] Yeah, we need to know what Watt looks like in the network. Like what kind of a...because there are a bunch of different algorithms you can run to find out the most influential people on there.

DH: [42:17] Though, they said... I have to say, Nayomi, this looks mind-melting.

NK: [42:24] I just tried to call up the graph although I don't know if that's the wisest thing to try to do live. We'll try to do it. Yeah, maybe go... Might just try to open it here. Yeah, so there's Stedman.

DH: [42:41] So this is Stedman's community?

NK: [42:42] Yes, Stedman's community, which is 1,323 or whatever it was. I do have the stats on how many people were in it and what time it spanned. So this is what it will... it gives me if I call up all people and in this case I've just brought in their associated locations. I see it's placed very centrally to the centre of that blob but I don't actually know what that means in terms of the way it's putting the graphs together. Because I see that it places people... that some people are on the periphery and it must be to do with the number of connections.

DH: [43:29] Could you bring in the instruments and guilds?

NK: [43:45] Let me just pull it out.

RH: [43:37] Also got the issue there of the difference between London and London W and I don't know how much of that is an artefact of a different period where you start getting postcode type. Or if it really stressed a to move to Westminster out of London or something like that.

AB: [43:55] I think we need to be able to view in both those... we need to sort of assign equivalent categories to earlier locations.

DH: [44:08] Pseudo postcodes are a little bit strange.

BJ: [44:14] It's difficult not... this thing of not knowing why. Because I've looked at Nicholl and Nicholl has got like a pretty average looking profile with an instrument maker which [inaudible]. Like I wouldn't be able to say this person looks amazingly important based on just looking at it. There's loads of people like them.

RH: [44:32] I presume it's... they have some apprentices and those apprentices... So their apprentices remain active.

BJ: [44:39] I mean Rust is also associated with James Watt. So this might be James Watt.

RH: [44:45] I'm not seeing Watt in the...

BJ: [44:46] No, I just noticed that.

AB: [44:49] I think if we draw the graph down, I suspect he will be off the slight tangent.

NK: [45:00] Yeah, I'll try. I'll try that when it spits it out.

BJ: [45:05] What were the other names?

AB: [45:18] How are we doing for time in terms of people's availability to...?

BJ: [45:21] I'm kind of okay actually, right now.

RH: [45:27] Yeah

BJ: [45:33] Can I ask about kind of timing and scale of this project, Nayomi? I mean, what are your ambitions for it? And how long does it last...this?

NK: [45:43] For the next two weeks! I've got a dissertation to get in on my work with the community detection.

BJ: [45:50] And is that working solely with SIMON? Or is that working with other datasets as well?

NK: [45:54] I've just worked solely with SIMON. I don't know if this is going to play ball. I've tried to bring in the guilds and the locations but it seems to be too much for it to manage this and Zoom at the same time. It will time out. I can hear the fan on my computer going. Yeah.

DH: [46:17] Yeah. Neo4j says no.

NK: [46:24] There you go. Too much for it. It died.

BJ: [46:26] Now this is all done in a browser? Oh, my God!

DH: [46:31] Have you got a server running underneath it?

NK: [46:34] There is. So it lets you set up a sandbox. I did try to download the database onto my computer, but then I had...there was...trying to integrate it with Python so that I could run some of the scripts from there. And it wouldn't allow me to connect. So I've done it through a sandbox. It seems to work quite well.

DH: [46:54] I see. Yeah, it seems good. Right. Interesting.

AB: [47:02] Can we see...I mean I don't know if there's anything else that you want to do...I'm keen to see the umbrellas.

NK: [47:07] Yes! I'll show you umbrellas. Let me get back here.

AB: [47:15] As one of those kinds of you know, weird categories...

NK: [47:18] It's just a manageable thing that I could show you because some of the communities are so large that it's very difficult...Yeah, here we go. So if I go to umbrella makers. So first of all I said, "Who made umbrellas in this network?' and it gave me this. So it gave me six individuals. And because I pulled out the community ID, I was then able to go in and have a look at what these communities comprise. It's quite interesting. So if I pull up...

RH: [48:01] What's the pinkish node?

NK: [48:04] Sorry, I didn't mean to do that.So the pink are the communities from Louvain... from the algorithm. Trying to find it again... here we go. Yeah, so the orange are the individuals and the pink are the algorithmically determined communities.

RH: [48:31] So they're all in distinct communities?

NK: [48:33] They're all communities. If then go into Pensotti...

RH: [48:41] They do seem to be our Italian [inaudible] types.

NK: [48:46] It is interesting. There are three distinctly Italian sounding names and three that don't sound Italian. So I was quite interested to see how the communities look different... and they did.

AB: [49:00] I've been doing some work on likely ethnicity based on stuff for name and surname. And I think that Franks and possibly Steers are likely the first or second generation as well.

NK: [49:12] That's interesting. So if I open up...This is Pensotti. Very small. In Dudley.

DH: [49:27] Umbrella makers in Dudley.

BJ: [49:31] Famous umbrella makers in Dudley. But why do we have Pensotti? I'm going to look them up now.

AB: [49:37] Because they'll be making instruments too.

BJ: [49:39] Well, of course. But what?

AB: [49:44] What would you speculate?

BJ: [49:46] I mean, it had to be barometers.

AB: [49:50] Is that right?

BJ: [49:51] Yeah, it's always barometers because that was the major spread of what we call scientific instruments, at least in our data in terms of retail was specifically Italian immigrant barometer makers, largely in Manchester, Liverpool I think but obviously in Dudley.

AB: [50:14] Portland too, Glasgow.

RH: [50:15] Specific gravity beads are a lot in Scotland. So glass because they come from mirror looking glass makers.

NK: [50:27] Sorry about this. I'm just trying to find a way to open the rest of them.

BJ: [50:43] So John Steer was an everything maker in Derby including "waitches", which must be watches. He made a fancy repository. I think that was probably the name of his shop.

NK: [50:57] Yes, that's what I got for Steer.

BJ: [51:01] He lived on Rotten Row [inaudible]. That's amazing. Okay.

DH: [51:09] Are they all Midlanders?

NK: [51:11] No, not all of them. I think Franks was a bit different.

BJ: [51:19] Sounds Manchester, doesn't it?

NK: [51.22] Let's have a look at that. That was more widespread. Yeah.

BJ: [51:25] I think Franks is quite a big family.

DH: [51:35] So interesting that they're in separate communities. So there's a slight... there's this one... like, is the instrument type serving as a bridge? I mean, I know it's just an artefact of the data. But it's not really a bridge between... but it sort of makes you think well, what not in the model that could be shaping the communities? I guess. So yeah, I mean.

RH: [52:17] In these... so we've got basically two surnames on this. There's the Franks group and then someone took over from Cook & Sons [inaudible]. Well, okay, more than two, but that's the main picture, I think.

DH: [52:35] This seems really interesting. They just yeah... like why are they all...?

RH: [52:42] There's also a few "See Also"s to worry about I think. Yes, I've noticed they've got the same surname.

DH: [52:52] But then there are "See Also"s but then it looks like with the Franks, they're all active in Manchester. So it seems sort of plausible that they might be the same family.

RH: [53:06] Good reason to "See Also".

AB: [53:09] Apart from the offshoot who moves up to Newcastle to make umbrellas.

DH: [53:16] So a wayward son.

BJ: [53:23] Optician, umbrella maker, coin seller and clothes seller, So he really...

AB: [53:27] This is our C Franks?

BJ: [53:35] Only known from the directory as that.

DH: [53:44] Then what we've seen... so B Franks took over from B Cooke who worked for Baurley, who was employed by T Cooke.

AB: [54:01] T Cooke was employed by Baurley. I think [inaudible].

NK: [54:07] So it should be the other way?

AB: [54:10] Well, it's... in a sense it's the right way. "Employed by" and "Worked for" are kind of synonymous rather than having... it's easy to misread them as being different things whereas actually they're the same thing. So one would imagine that Baurley is probably the starting point for this, who employs two Cookes. No, no, that's wrong, isn't it? Sorry.

BJ: [54:45] That is wrong. It's a little bit confusing. I actually can't find the relationship between Baurley and Cooke in the data but obviously it's there as one...

NK: [54:54] And we could see it, I suppose, if we tried the evolution on this. Growing the thing.

AB: [55:02] Yep, that would be interesting to see.

NK: [55:03] Yeah I might try that on this particular community.

AB: [55:09] Yeah, that could be very interesting.

BJ: [55:15] Oh, I see. I think has to do with the relationship. So Baurley worked for Cooke & Son but it doesn't specify more than that and so T Cooke & Sons and B Cook & Son who both seemingly appeared as a relationship when there's only actually one of them specified for Baurley.

DH: [55:41] The relation might be on T Cooke.

BJ: [55:45] One of them will be on T Cooke and the other one will be a disambiguated relationship from Baurley to B Cooke.

DH: [55:53] Yes. Yeah.

BJ: [55:56] A bit of a nightmare. So might not even be correct.

NK: [56:01] Oh, I see. So this might not be...

BJ: [56:06] Yeah, that one has been as one where someone's disambiguated it and they've said that's B Cooke but that might be because it's & Sons, so there's some evidence for that. But then there's ever contra evidence in the T Cooke & Sons record.

DH: [56:22] Yeah, but B Cooke and T Cooke are definitely related with one another, aren't they? Just looking at the IDs, they've been added to the database sequentially, haven't they?

AB: [56:36] Sorry, just Boris quickly: are you saying that B Cooke & Son has been possibly disambiguated as B Cooke?

BJ: [56:48] No, I'm saying that in the record for Baurley or Bowerly, it says that he was an employee of... he worked for Cooke & Son, phrased like that. And that's the only way that I can see that he could have ended up linked to B Cooke & Son. B Cooke & Son there's no mention of Baurley. But obviously if you go to T Cooke & Sons, I would wager that...

AB: [57:15] Okay. Okay.

BJ: [57:16] So probably a false disambiguation there.

RH: [57:29] The other thing that looking at this would be interesting to bring into some of is sources. So we feel we're characterising this as the results of looking through trade directories, which obviously gives you different kinds of information than some of the other sources. And the database throughout. Obviously, over time, you've got different types of sources that are likely to be the major ones for different periods, or different localities. I'm not quite sure whether that would help or where I'm going with that. But I think it would tell us something about why certain connections are made or why some people end up not being connected.

BJ: [58:15] Yeah, I mean, just being very kind of simple minded about it, I would say that Stedman is the key to all of this because that's one that none of us in a million years would have predicted as the most well-connected person. So, my immediate question is what is going on? This is what, when we found out that W Harris was very well connected, we kind of plumbed into Harris and found out quite a lot about him. We just haven't happened to have done that with Stedman. And we certainly... even with Harris, we went straight into looking at the sources and doing it in an old fashioned way, whereas doing it computationally with Stedman would be just really interesting because I have no idea what that means. It kind of feels more like discovery-like than... These other things are very... they seem to be very rich tools for that kind of investigation.

AB: [59:19] Yeah, I mean, there is a very... we have a very clear question that we want to ask about Stedman. And we can facet the data around him in ways to try to understand that.

BJ: [59:31] Yeah. I mean, it's an interesting question because it's not even clear how you would...maybe Becky's got better ideas. I don't even know exactly how I would go about trying to answer it. It feels like a question you can only answer computationally because the evidence itself is statistical. So I could kind of make a biographical narrative about why he might be important but I want actually try to find out why it thinks he's important and what that looks like.

RH: [1:00:05] Yeah. Is it about him? Or is it really about his connections? Maybe two generations away, possibly.

DH: [1:00:13] Yeah, that's really interesting.

AB: [1:00:19] He's emerged from an analysis of this kind of the entire dataset. It would be useful to segment the dataset in other ways and try to kind of surface other individuals in that way, for example, in provincial centres or for particular periods or in particular areas of London, for example, to understand who are the dominant figures in one way or another more locally.

DH: [1:00:55] Nayomi, have you got any PhD plans?

NK: [1:01:00] I would love to. This seems to me to span a PhD [inaudible]...it's fascinating.

BJ: [1:01:17] Now, it's really nice to see. I mean, this is much richer and more complex than anything I was able to do in Gephi. And I've just had to go at doing that kind of significance. But I would then... the way I did it was just by grading the nodes by size based on a few of the different significance algorithms. But it wasn't... it didn't kind of do enough to differentiate. And so I ended up being a little bit frustrated with it.

DH: [1:01:46] Seems quite, it does seem quite interesting because I'm speculating that... I mean, as you said, Becky, that perhaps, Stedman, within his own generation, perhaps he's got this artefact that we know a lot about his apprentices. But then, maybe he's connected to some very, very highly connected people or a lot of people who have... they're not coming up high enough to get a

high PageRank algorithm but in aggregate, they're giving him this really high score.

AB: [1:02:24] And how is he useful to them, if that is the case? How is useful to a collection of individuals of that kind who are perhaps more intrinsically influential than he is? I think it would be fascinating to continue this dialogue with this method and to be able to take subsets of filtered data like this one that we're looking at right now back into Gephi or into a geospatial analysis and develop a more nuanced understanding... see whether it's possible to develop a more nuanced understanding around questions like that one about Stedman. Would you agree with that, Boris and Becky?

BJ: [1:03:08] Yeah, certainly. I mean.

RH: [1:03:10] It's always interesting. Yeah, we can find more Stedmans that would certainly give us...

NK: [1:03:17] Something that this sort of software is quite good at doing is looking at "hops away". So, you can quite easily look at Stedman and individuals who are one hop away from Stedman, two hops, three hops. So that's quite easy to do with this sort of software.

AB: [1:03:37] Yeah, I think it'd be very... I mean... I'd like to continue the conversation, Nayomi. You probably just need to concentrate on getting your dissertation finished but maybe we could speak again either before or soon after that?

NK: [1:03:53] Yep. Well, more than happy to... I feel like I could play with this forever.

DH: [1:04:00] Great. Nayomi, we should stay on the call just to sort of wrap up but thank you so much everybody for your time.

BJ: [1:04:12] Well, thank you. Thanks for sharing it with us. It's really interesting work. And I hope that the next two weeks aren't too hectic.

NK: [1:04:22] You've all been there!

AB: [1:04:25] Good luck with it.

BJ: [1:04:27] Okay.

NK: [1:04:28] Thank you for your time.

DH: [1:05:30] Thank you very much. Bye, bye.