This is a transcript of Video 4 in William I. Wolff's webtext "Twitter Archives: A Discussion of Systems, Methods, Visualizations, and Ethics," published in Kairos: Rhetoric, Technology, Pedagogy, 22.2, available at http://praxis.technorhetoric.net/tiki-index.php?page=PraxisWiki%3A_%3ATwitter+Archives

Hello! in this tutorial I'm going to discuss moving from DMI-TCAT through a Co-hashtag export into Gephi to create a co-hashtag network. To complete this you're going to need your DMIT-CAT analysis page open, if you don't recall where that is you can go to the Word document or other file where you have saved your login information and that will be this URL and then this username and this password and you'll also need to have Gephi open and of course you'll need to have your DMI-TCAT. For this one, I'm going to be archiving the Terrence Crutcher hashtag, this will work best with when you are archiving hashtags, I guess it could work for anything, but you know archiving hashtags, it really works very well.

I've got myself a good number of tweets about 500,000 little 580,000 I've only chosen a day's worth of tweets and you can choose something much larger if you'd like, you won't want to get too much smaller than that if you can if you can help it. Then you would scroll down to networks, and we're going to do a co-hashtag graph. And what a co-hashtag graph does, is it maps relationships between hashtags so any time to hashtags have been mentioned in the same tweet there is a edge connected to it and the hashtags become the nodes. This is an undirected graph and this is nice where you can locate issues that are related to it you can see how people are using similar hashtags together and so on.

There are two options to set the minimum frequency which is how often the hashtag is used or to get the top hash tags and for this for our purposes we're going to be using the top hash tags so just click launch and the default here is 500 which is much more than we really need so for our purposes I'd like you to choose 50. Okay. And then click OK. now if you find as you're going through this that 50 is not giving you the results that are really meaningful, then you can up it to 100 or 200, 500 is really a lot. I'll click ok and it takes some time to download so I've already gone through that process and we're not going to sit and wait for that so you could do that on your own time. Ok, so let's go over to Gephi and we've opened this up and we could open a new project, open, and I'm going to choose the top 50 file that I've downloaded. Open.

Now when you open this, oftentimes you will get this error. This is an error that I've notified the DMI-TCAT about I guess Gephi should be notified of it as well, it's okay, but you what you want to make sure that you do is you write down this number. Whatever this number is please just jot it down because we're going to use it in a minute. Notice here that this is an undirected graph which is what we expect, there are 7600 nodes with 355 connections. So even though that we even though we have exported the top 50 hashtags, we still get 7600 nodes, I'm not completely sure why that is I've asked them a question haven't received a response just yet, but these are the connections and I imagine in what's happening is it's giving us all the hashtag but it's only giving us connections between the top 50 hashtags or what those top 50 hashtags are but I'm not completely sure. So click OK and you will get our Borg-like graph, and so what we need to do now is we need to go into the data laboratory, ok, and you remember the number that you just wrote down I just wrote down 6275 so I need to scroll down to 6,275. And, we notice that, the ID number for our the hashtag that we have been tracking the Terrence Crutcher

hashtag, and we'll see that it's missing this word frequency multiplied by unique users integer. And that's because the number that is calculated that is greater than the maximum number that Gephi allows.

And the way that is calculated is you take the distinct number of users which is 311,931, and multiply it by word frequency… 584 821… Okay, so there are the two this total number of distinct users using that hashtag and then there are it is used, in terms of, throughout the archive 584 thousand times… Now, if I were going to calculate that which I have done the answer is 182 billion 423 million 799 thousand three hundred and fifty one. That is distinct number of users multiplied by word frequency. And, if you have gotten an error where you have not calculated, or Gephi, did not calculate that number for you, in order to fix the error you need to do those calculations…

So, you take word frequency and multiply it by distinct number of users. Okay, now I get 182 billion 423 million 799 thousand 351. I try to click enter, and it gives me this error message because that number is too big for Gephi. The largest number that Gephi will allow is 999 million… 999,999. 999 million 999,999… so if you get an error when your calculations come through, you can just delete the number and enter in the maximum 999 999 999… Now what might happen is that you get two errors, like the two of the hashtags are thrown back in there… If that is the case, then I would recommend putting in for the one that you know is use the most, the larger 9 9 9 and 9 and then just putting out a smaller one like 777 million seven- something like that- so you can give a different range if you want to be more precise, you can sort of figure out the percentage differences between the two and then calculate it. But for our purposes, that is really is not necessary.

So once you have done that, you can go back to the overview… and we've got our Borg area over here, oops, I'm going to shrink this down, and we were going to basically run the similar kind of statistics and filters as we ran in the past. So, for the first one, we're going to take a giant component filters area drag it down, and click filter. And, we can see that it reduced it considerably actually we now have from 76 06 we get to 51 but all the edges are connected so what that's telling me is that these top nodes are the most important ones because we export top 50 it's given us 51 of them, all the other ones were disconnected, and just sort of superfluous, next what we're going to do is we're going to get our degree range why you might not need it this time around as we have in the past but if you want to filter the degree range you can but it will give us something very, very small.

So, I would just sort of keep it the way it is. Then we can go and we can choose our layout which again would be the Yifan Hu, and it squishes everything down very small so we could change this to 500, and it will bring things out for us so that we can start to see what is happening inside, inside of here. Then we go over to our statistics, and we can run the modularity again. And for modularity with something this small, you might need to go below one. So if I ran one this first time, you can see I only get three communities, which really is not enough to learn anything interesting- so I'm going to want to run this again and I'll do .5 and we'll see how many we get and here we get five communities and that could be enough for a small map like we're, like we're, doing here and we don't have to do the network diameter, because these are just direct

connections, one to the next to the next this is a co-hashtag graph so there are lines connecting to two dots so that's really all we need to do, uh- for, for this.

And once we've completed that, we go over to the area to the top left. Now again, you might have, mine says partition, uh you might have something that says attribute, so click on attribute, because it says choose an attribute, I would choose modularity class and you can see I have my five communities, then I click apply and you can see that they're sort of starting to break down, into pretty clear distinct areas and then I will go over to ranking, I'm sorry, I would click on little circles over here, click on ranking, choose an attribute, I'm going to choose degree and I've chosen my minimum of 10, maximum of 50, click apply and I guess I can run my "no overlap" just in case anything is overlapping just a little bit here…another one overlapping there that I might move out of the way just a little bit… okay, and that is pretty much it.

Then I would go to the preview, and refresh and get this crazy thing, click on rescale weight, and now I see the more lovely one I'll expand my lines a little bit, refresh that, get some nice edges I got a little overlapping here which I'm not thrilled about, I might go back and fix that and then I'll show my labels, don't want the portion size and I want to choose this font that I like, I'll right click refresh…. hey sweetie, how ya doing?... Ok, so now I've added my the fonts that I like, and I can go back and take a look and see, and start looking through, and see what is happening at, at these uh, at these nodes… and I can start thinking about why, what's interesting about these, as I start to see that if I'm starting to think about the story of this hashtag is that I've got the "say their names" hashtag and then most of the nodes are the names of men or women who have been killed by, by police killed by police.

My son joined the room and I probably should not have said that so we see that these were all connected, interesting that they are also connected to, like Brangelina, so I'd have to go back and take a look at the tweets to see why those are, why those are there. I can then just go and I can export it, as a file and then you can upload those files or use them as that as they want. Ok, so again this was a co-hashtag graph from DMIT TCAT to, to Gephi, and if you have any questions, let me know.