Studying 'fake news'

Hello, today I'd like to talk about studying fake news. And I want to start off with the research question or a kind of generic one. And that would be to what extent is fake news, which we could also refer to as misinformation resonate in political spaces across social media in the run up to whichever national elections?

Now this is a generic question. Oftentimes fake news is studied in the run up to national or other elections. But what I'm going to do today is I'm going to talk a little bit more specifically about one in particular, and that is the the presidential elections in the US in 2020. So, this is the sort of reformulated more specific research question to what extent is fake news resonate in political spaces across social media, in the run up to the presidential elections in the US?

Now? I've asterisked three terms there. And we're going to go into them one by one. The starting from the last one, the president presidential elections, and then we'll talk about political spaces and how to demarcate them or, what one can do in order to demarcate though, and then the tricky term, fake news and the fact that it is a, it's a classification problem and it's also one that is politicised, so we'll talk about that.

So, first, as I said, oftentimes, so called fake news is studied in relation to elections. Of course, this is because of the sort of birth of the fake news crisis, which was in connection with the 2016 US presidential elections. And this birth, you could argue, was from a particular news article that was written by Craig Silverman that appeared in November of 2016. With this quite shocking graphic, where the headline read this analysis shows how viral fake election news stories outperformed real news on Facebook. Now, there are a couple of studies that he had done with colleagues. Prior to it, he was the one who discovered the Macedonian fake news farm, and which was producing first initially stories that favoured Hillary Clinton and Bernie Sanders, but they weren't earning enough money. So, they switched to Trump. And the stories that they were putting out, got quite a lot of engagements.

There was a subsequent story and then and then, and then it's one in November is when he actually used the term fake news. And in the story used to fake news, the term also rather interchangeably with false news. But nevertheless, this this, this, this graphic, and the headline is, in some ways, the sort of birth of the fake, fake news crisis and that is the crisis being that it's outperforms mainstream news, and it did so, right before the 2016 elections and this is on Facebook, of course.

Now, what is interesting about this particular article, and this is the definition that they use for fake news, and it includes sort of imposter sites. So, sites or pages that are purporting to be real news organisations, which are not.

There are now and again considered some sort of conspiracy sources, but mainly the other one is hyper partisan. And this, and this is an this is an interesting term, which will which we'll break down in a

minute, but what Silverman did also with his colleagues is it did a qualitative determination. So, they looked at the stories and from a journalistic kind of fact, using fact checking eye, and they made this they made this determination.

So and then subsequently, there are many so called fake news studies that were done prior to European elections and sort of Sweden, Germany, France, the Netherlands. So, these were, this is the sort of typical subject matter, but also there's a lot of misinformation studies, requests done around the pandemic or the so called infodemick. And around other sort of health information, etc. So So, elections and pandemics, so to speak, or those are the favourite sort of rather bounded events, bounded and sort of timeframe. And then you can do like this time series analysis. So, is it getting paused? Is the problem getting worse? Is it improving? etc.? Is Facebook doing enough?

Okay, so then the second aspect is, in the research question about the extent to which fake news resonates in political spaces across social media in the run up to a national election. So, what are these political spaces? Well, political spaces are oftentimes, and this is a term that sometimes used, but you can also drop the term, but what you're doing is, is you are creating a set. So, you're querying, you're building a list of, of queries, and then you're querying them in the platforms, and you're getting results back. And these results are then sets, source sets, which are oftentimes ranked by engagement. And then you're, you're checking these sources for misinformation, or quote, unquote, fake news.

So, this is the, these are these, these are these political spaces. And you can build your lists and query across social media platforms from the public facing ones, to the more anonymous or masked ones like 4chan, or Reddit. And then and then get some sort of sense of the extent to which the misinformation problem is across all platforms, or whether it's more a problem of one or two, etc.

So, here today, you see that Facebook, there is bolded, I'm going to talk specifically about Facebook, and not go through all these all these other platforms. But at the end, there's a bibliography where you can where there are resources, where we've studied all these other platforms, as well, and you can [inaudible] will be there.

Okay, so how do you do this? Well, when you build a list of prior to an election, the straightforward parts of the lists are the are the politicians names, normally want the party leaders or the main candidates, the party names. And then you want to make a list as well of election issues. You could also add the name of the election as well, I haven't put that on here, but you can add the name of the election. Oftentimes, if you're querying the reasons on or if you're querying Twitter, and you're doing you're querying hashtags, and there's normally one or two dominant hashtags for a particular election. But here, it's the list of politicians' names, the party names, as well as election issues.

Now you can do the election issues in a couple of different ways. Sort of the most straightforward way is to make a generic issue list. And you see that here, for in the UK for YouGov. And you see that these are quite generic names, and you can use those terms to query. But when you do, you probably want to also add the word election. Because otherwise, those terms are too broad. Or you can make, you can do a little bit more sophisticated manner and make and use the party platforms. So, on the screen, you'll notice that there's a sort of issue list from the Democratic Party in the US. And you could see that

they have quite specific terminology. So, then you're using party related issue language. So here, it's slightly more complicated, because you'll create an issue lists per party sometimes overlap, sometimes they won't, you probably want to retain the specificity per party and query all of the terms to see whether or not the misinformation problem, if there is one is related to a particular party far more than other parties or whether it's across the board.

Now, once you have your list of politicians names to the party names, as well as the as well as the issues, how were you going to do it, then there are a couple of ways for gathering Facebook data. So, these are two of them. These uses or repurpose tools that have been developed for different reasons for marketing, largely marketing research, although the crowdtangle too is now being used by routinely being used by academic researchers. So, there's crowdtangle. And here you can, you can study which Facebook pages are resonating. Right, so which way you query and then you see which equated see, which Facebook pages are engaged with. And you see the example here, I put in social security, you don't have to use quotation marks that there are for exact searches. And you have a date range there, and you're, you're returning the data by total interaction.

So, this is the engagement score. So, it'll be a ranked list of those pages that like, in some ways, what Craig Silverman did, those pages that are being engaged with or Silverman, in fact, used buzzsumo, which is a marketing tool. And when you query that, you will use the quotation marks, you'll there are also a couple of other settings you can, there's for a particular region, region setting, there's also a language setting that you want to use. And then once you get what you are returned to your web URLs. And in fact, this is how we did the study we used buzzsumo. And you get the query here is climate change, and you get a sort of ranked list of URLs. So, with the with the date of the post, the host, but then beneath that has a clickable link to the story, which is important, which we'll come to. And then the engagement scores are the number of times it's been liked, shared, commented upon, or reacted upon more broadly.

Okay, the third part of the research question is fake news, or misinformation now, this is a as I said, it's a it's a difficult classification problem. It's also politicised. So, I want to take you first to Silverman's sort of idea from BuzzFeed news idea of fake news. So he basically had two three, alluded to a fourth definition so that so the first type of quote, unquote, fake news source is, what do you referred to as sort of imposters or sites that purport to be a news organisation when in fact, they're not. So that's number one.

The second one is to use this term, hyper partisan. So, the term itself was put into play by John Harmon of the New York Times, couple of months before Silverman's study. So, he used this term hyper partisan, he also was connected to Facebook how, quote unquote, hyper partisan sources. So, sources with sort of clear ideology, are prone to hyperbole were doing rather well on Facebook.

And then the second one, or the third one, uses the term misinformation. That's why I have it in parentheses and the research question, he doesn't explicitly refer to conspiracy, although in related articles that does come up. But nevertheless, if you look at the top five, sort of, quote unquote, fake news, election stories, you see that the top four are from one type, these are the quote unquote, hyper partisan and then the fifth one FBI agents suspected in Hillary email leaks found dead in apartment

murder suicide, Denver Guardian. Denver Guardian sounds like a real news organisation, however, it's not an imposter. And so he talks about these two types of in particular, and then these are the examples the other ones the most, the one that resonated the bosses that is the Pope endorsing Trump, which was on this Facebook page, which also has a URL, but this is for the Facebook page. Which is, which he discusses as this remarkably successful, utterly untrustworthy site, which does not publicly list an owner or an editor. And this is an important piece of information here, because then it gets you into how journalists, oftentimes evaluate sources do particularly source criticism on the basis of journalistic standards.

Well, before we go there, I just want to talk a little bit. So, this is this is Silverman's definition, this is the broad definition of fake news. And as I said, he uses, it defines them as both an imposter as well as hyper partisan, sometimes alluding to conspiracy. I've also added here disinformation, which is oftentimes part of this larger definition, like, like Russian, IRA influence, campaigning, this sort of thing. So, this is, and so that's where the line is drawn, so to speak, when beginning to classify different kinds of sources.

So, this is one way of looking at this fake news. Now, another way is if you move this line up, and do not include hyper partisan sources as part of this definition of fake news and misinformation, and this is in fact, what Facebook does, when in their definition, what they call false news. But also, some journalists, so the one that I'm referring to in the published studies, the Dutch quality newspaper called the NRC handles blots. And when they did their own sort of evaluation of stories, and the run up to the to the Dutch elections, they didn't include hyper partisan or tabloid or these sorts of sources as being quote unquote, fake or misinformation, rather than the their definition was quite strict, like, like, like Facebook. So, these are, these are the two.

Now I want to begin with, how journalistic organisations and media organisations begin to think about sources as being problematic or not. So, what you see here is a tool called news guard, that's also a plugin you can install on in your browser. And then when you go to a particular source, like the two here christianitytoday.com or oann.com, it gives you a quote unquote, nutrition label. And it's about the extent to which they meet particular journalistic standards. So, you see, Christianity Today meets a lot of them, whereas the, the other sourced Does, does not. So, you see how sort of problematic information, misinformation, fake news begins to be addressed as something to be classified by news by news organisations.

Now, I want to talk a little bit more specifically about this notion of hyper partisan. And so as I mentioned before, and as you saw, it was this defined by John Herman has opened the ideological web operations of both ideological on the one hand, but then also this internet or web phenomenon on the on the other. So, there are a number of different, let's call them media monitoring sites, which classifies sources, and they classify them according to quote, unquote, their neutrality. So how neutral are they and then, and then they put them on a political spectrum. And you see the media bias chart, it's also oftentimes referred to just as the chart, one of them where they also have picked up this term, hyper partisan now this one was published in 2018.

So you see that there's sort of something of a continuum here, where particular sources are classified and there's one here in the UK This fightingfake.org.uk, which also sort of classifies sources as such, now know that the one use of the term hyper partisan, the other one doesn't, but the ones that are the on the farthest ends, so to speak, and so from the media bias chart, that those would be the two sides of the hyper partisan the most extreme, and then that's on both sides of the political spectrum. And then for the fighting fake, it would be just the left or the right would be considered more partisan. Although you could also think of some of them here as hyper partisan. If you if you look at this, if you look at the URLs,

And this is how it's done. So, there's a sort of scoring system to this is this media bias fact check, which is another one of the sort of media monitoring organisations that that then classifies them, according to political, political leaning. Now, what's interesting is that one can make use of these to do this sort of work. But I want to, before we make use of that, I want to point out something interesting. So you look on the all sides and you notice that on the left, sort of the light blue, you see the New York Times and then the farther on the left the dark blue, you see New York Times opinion. Okay, so what this alerts you to is the fact that you probably want to do your classification on a story level, as opposed to a source level, because it differentiates in this particular case between the news and the, and the editorial as being slightly different. So, this is the first point. So, the New York Times is in this by all sides is not hyper partisan. But it's according to all sides their opinion is, so that's, that's number one.

So, but you we can consider triangulating these. So, so I mean, these are all considered of decent quality, there are some issues with all of all of them. But nevertheless, could consider using all of these sources. And then if there's disagreement using the majority classification. So, this is this is so we're, we're in some ways, outsourcing the classification to, to these media monitoring organisations, who themselves profess different levels of neutrality.

Now. So what we do then is, you have you, you had your lists, you queried them in the software, you got back the results, in this case, web URLs, they are ranked by engagement score, and now we're going to classify them. And we're going to classify them in two ways. So, they're going to be classified as sort of mainstream or problematic and problematic is, is sort of what we're talking about is as fake news or misinformation that is to say, in in that sort of narrower definition. So, imposter, disinformation, conspiracy, this would be that. So here you have businessinsider.com motherjones.com, USA Today. So now, none of those here are considered problematic. They're all considered mainstream. But we have another, one another column, where we are considering whether they are hyper partisan or even partisan, so on here well, hyper partisan liberal, most extreme liberal, not the skews, but the dark. So the dark blue L and the dark red R and then from all sides and then the most extreme liberal hyper partisan liberal and then live partisan conservative, most extreme conservative so Mother Jones, in fact, you see here is hyper partisan liberal. So, we put it there. As we you just use the term progressive, you could also use the term liberal. So, this is, so there's two, there's two classifications, the main classification and a sub classification.

So now we're ready for the analysis. And so, let's return to the research question what extent is fake news and misinformation resonant in political spaces? So, this will now be Facebook, in the run up to the presidential elections in the US in 2020. So, we can look at, let's say the amount of it, right. But this

amount, we can also compare to what Silverman found in his in his work, so we'll do it, we'll do the comparison by proportions, just very straightforward way.

The second thing is because we had this differentiation, these two classifications, so the one being mainstream versus problematic. The other one is whether it's hyper partisan or not. We can also look at the difference definitions, the scale of the problem according to a narrower and a broader definition of quote, unquote, fake news. And then, finally, we can look into its politicisation by examining the extent to which the hyper partisan sources, which could be classified in the broader definition as quote unquote, fake news, whether they tend to be more conservative or tend to be more progressive. So, we can see that in some ways, the politics of fake news classification used in this third one, okay, so let's look now.

So, the data here, we're just running data from the first quarter 2020, and comparing it then to the first quarter in 2016. And we have very simple proportions that we're using. And what's interesting is that this is then the broad definition. So, Silverman's definitions, we're including in our in this particular one, the, the hyper partisan. And you see that that, in fact, the problem, the misinformation problem, or the fake news problem on Facebook, as worsened slightly. So, there's a bit more of it, then then there was in the first quarter of 2016, in the run up to the end of the elections. However, if you compare the broad definition with the with the, with the narrower one, so there's sort of false news, then you're not including hyper partisan sources, it's the problem is, is much smaller. So, whereas it was one in three and a half, now, it's one in nine, so that the problem gets reduced quite a lot, which I guess you would imagine but nevertheless, it's quite stark.

Now, moving to the question of the politics, here, we see that when the hyper, that hyper partisan, conservative sources are far more present in the in those that have been most engaged with, then the progressive ones. So, you so you see the politics of fake news classification quite starkly. So if we, if we use the broad definition, and we count hyper partisan sources as quote unquote, vehicle misinformation, then what you're basically doing is you're labelling a lot of conservative sources as that and far fewer progressive ones as such.

So in terms of the implications, so for the, if we have the, if we have the broad definition, the problem is quite large, which might imply that it would be addressed by automated means or that that might, one might think that it's the problem can't be addressed. But if it's we do the narrower definition. The problem gets smaller, it's still not that small. But nevertheless, one could begin to consider different forms of fact checking our sort of crowdsourcing solutions, quote, unquote, okay.

So, this is the first implication. But the other one is what I brought up here at the end, and that is thinking about the larger question of the politics of demarcation. So, you know, if we, if we're classify and call, a lot of the hyper partisan conservative sources as fake Then you you're politicising, and you will likely, as we have seen, over the past few years get a particular form of political backlash.

So here are the works of reference, Craig Silverman's work the famous sort of birth of the fake news crisis article, John Herman's coinage of the term, hyper, hyper partisan. And then the further readings

where you can dig in deeper to how to study them across multiple platforms and do sort of a kind of fake news study of with a cross platform approach. Okay, well, thank you very much.