NDI

تحليلات البيانات لمراقبة وسائل التواصل الاجتماعي

دليل توجيهي بشأن تقنيات الرصد والتحليل في وسائل التواصل الاجتماعي والأدوات والمنهجيات

May 2020

NDI.ORG

محتويات

المؤلفون ____ 3

نبذة عن المعهد الديمقراطي الوطني للشؤون الدولية (إن دي أى) ____ 3

شـكـــر وتنـويـه ____ 4

المـقـدمـة ____ 5

المعلومات العامة ____ 6

العمل على جمع البيانات ____ 8

تحليل البيانات والشبكات ____ 13

تحديد المؤثرين والمجموعات والحسابات ____ 20

تحليل الحسابات والمحتوى ____ 25

الـــخـــاتـــمة ____ 31

الملحق 1: مثال على رمز واجهة برمجة التطبيقات - جمع البيانات من واجهات برمجة تطبيقات البحث والتدفق من Twitter باستخدام حزمة إعادة التغريد ____ 32

الملحق الثاني: أدوات OSINT ____ 34

الـــمـــراجـــع ____ 35

المؤلفون

نيك موناكو هو مدير مختبر الذكاء الرقمي في معهد المستقبل وهو خبير في المعلومات المضللة على شبكة الإنترنت واستخدام البوتات السياسية، وهو متخصص في التضليل الصيني. وفي سياق عمله تشاور مع صنّاع السياسات وخبراء التكنولوجيا في كل من الحكومة والصناعة بشأن أفضل السبل لمكافحة التضليل والحفاظ على نزاهة الانتخابات في البلدان في مختلف أنحاء العالم.وقد عمل سابقا على معالجة هذه القضايا في شركة جرافيكا، و هي شركة تحليل الشبكات الإجتماعية و تحليل التهديد الإستخباراتي، و مؤسسة جوجل لتحليل الحقوق الرقمية، جيغساو وهو أيضا شركة بحثية تابعة لمشروع الدعاية الحاسوبية (ComProp) في معهد أوكسفورد للإنترنت.

دانييل أرنادو مستشار في معهد الديمقراطي الوطني لاستراتيجيات المعلومات ,ويغطي التقاطع بين الديمقراطية والتكنولوجيا, مع تحمل مسؤولية خاصة تتمثّل في تطوير برامج تتبع المعلومات الكاذبة وتعزيز سلامة المعلومات في مختلف أنحاء العالم .وفي نفس الوقت، يعمل زميلاً في مجال الأمن السيبراني في كلية جاكسون للدراسات الدولية بجامعة واشنطن حيث عمل في مشاريع في البرازيل و ميانمار والولايات المتحدة.

كما تعاون مؤخرًا، مع مشروع الدعاية الحسابيه التابع لمعهد أكسفورد للأنترنت وتركز أبحاثه على الحملات السياسية عبر الإنترنت والحقوق الرقمية، والأمن الالكتروني وتكنولوجيا المعلومات والاتصالات من أجل التنمية.

نبذة عن المعهد الديمقراطي الوطني للشؤون الدولية (إن دي أى)

المعهد الديمقراطي الوطني للشؤون الدولية عبارة عن منظمة غير حكومية غير ربحية وغير حزبية تستجيب لتطلعات الناس في مختلف أنحاء العالم للعيش في مجتمعات ديمقراطية تعترف بحقوق الإنسان الأساسية وتروج لها.

ومنذ فتح المعهد الديمقراطي الوطني أبوابه في عام 1983، عمل المعهد و شركاؤه المحليون على دعم وتعزيز المؤسسات والممارسات الديمقراطية من خلال تعزيز الأحزاب السياسية والمنظمات المدنية والبرلمانات، وحماية الانتخابات، وتشجيع مشاركة المواطنين والانفتاح والمساءلة في الحكومة.

ومع وجود موظفين وممارسين سياسيين متطوعين من أكثر من 100 دولة، يجمع المعهد الديمقراطي الوطني بين أفراد وجماعات لتبادل الأفكار والمعرفة والخبرات. ويتعرض الشركاء على نطاق واسع لأفضل الممارسات في التنمية الديمقراطية الدولية التي يمكن تكييفها مع احتياجات بلدانهم ويعزز نهج (إن دي أى) المتعدد الجنسيات حقيقة مفادها أنه على الرغم من عدم وجود نموذج ديمقراطي واحد، إلّا أن بعض المبادئ الأساسية تتقاسمها جميع الديمقراطيات.

ويتقيّد عمل المعهد بالمبادئ المنصوص عليها في الإعلان العالمي لحقوق الإنسان كما أنه يشجع على تطوير قنوات اتصال مؤسسية بين المواطنين والمؤسسات السياسية والمسؤولين المنتخبين، ويعزز قدرتهم على تحسين نوعية الحياة لجميع المواطنين لمزيد من المعلومات حول المعهد الديمقراطي الوطني يرجي زيارة الموقع
www.ndi.org.

حقوق النشر
© المعهد الوطني الديمقراطي (NDI)
موقع الويب www.ndi.org :

حقوق النشر © المعهد الديمقراطي الوطني للشؤون الدولية جميع الحقوق محفوظة يمكن إعادة إنتاج و/أو ترجمة أجزاء من هذا العمل لغير الأغراض التجارية يتم الإقرار للأغراض التي تم الحصول عليها من قبل إذن كتابي مقدمة من المعهد الديمقراطي الوطني ويتم الاقرار مصدره المعهد الديمقراطي الوطني يتم إرسال نسخ من أي ترجمة إرسال طلبات إذن النشر إلى موقع legal@ndi.org

شـكـــر وتنـويـه

يعرب المعهد الوطني الديمقراطي عن امتنانه للوقف الوطني للديمقراطية(NED) لدعمه لوضع هذا الدليل. إن الحوار الوقف الوطني عبارة عن مؤسسة خاصة غير ربحية مكرسة لنمو وتعزيز المؤسسات الديمقراطية في مختلف أنحاء العالم كل عام وتقدم الشركة أكثر من 1600 منحة لدعم مشاريع الجماعات غير الحكومية في الخارج التي تعمل لتحقيق أهداف ديمقراطية في أكثر من 90 بلدا منذ تأسيسها عام 1983، ظلت الأوقاف على رأس قائمة الصراعات الديمقراطية في كل مكان، بينما تطورت إلى مؤسسة متعددة الأوجه تشكل مركزاً للنشاط الموارد وتبادل الآراء بين الناشطين والممارسين والباحثين في مجال الديمقراطية في جميع أنحاء العالم ,كما نشكر ايضا معهد المستقبل لتعاونه وشركته وتواجيهاته ان معهد المستقبل ليس ربحية ومخصص للمستقبل المدني ان معهد مستقبل مخصص لاضخم الموئسسات الأجلة علي مستوي العالم.

فعلى مدى أكثر من خمسين عاما، كانت الشركات والحكومات ومنظمات التأثير الاجتماعي تعتمد على التوقعات العالمية لمعهد المستقبل، والبحوث المخصصة، والتدريب على البصيرة والمبالغة في الإبحار عبر التغيرات المعقدة وتطوير استراتيجيات جاهزة للعالم وتوفر منهجيات المعهد ومجموعات الأدوات آراء متماسكة حول التحويل. إن الاحتمالات التحويلية في مختلف القطاعات التي تدعم معاً مستقبلاً أكثر استدامة

التصميم والطباعة ،إرون مارك:2020

المـقــدمــة

أصبحت وسائل الإعلام الاجتماعية جزءاً متزايد الأهمية من المحادثات التي يشارك فيها المواطنون والمرشحون والأحزاب والمنظمات ذات الصلة في الأحداث السياسية والانتخابات والاستفتاءات، فضلاً عن التصويت على مشاريع القوانين والإضرابات وغير ذلك من أشكال النشاط السياسي من المهم بالنسبة للباحثين، ومراقبي الانتخابات، ومنظمات المجتمع المدني، والأشخاص العاديين أن يجهدوا أنفسهم بالأدوات والأساليب والممارسات اللازمة للمساعدة في جمع وتحليل البيانات من الفضاء على الإنترنت ويشارك أعضاء منظمات المجتمع المدني الدولية، بما في ذلك الباحثون، ومدارء البرامج، والناشطين وغيرهم، في برامج مختلفة على الصعيد الدولي، بما في ذلك العمل كأجزاء من بعثات مراقبة الانتخابات، لدعم المجموعات المحلية لتطوير قدراتها على الرصد، أو الرصد على خطاب الكراهية، والاتجاهات السياسية، ومجموعة من المواضيع الأخرى.

وهذا دليل لمساعدة الباحثين والمراقبين في الانتخابات، وخبراء التكنولوجيا وغيرهم على فهم أفضل الممارسات والأدوات والمنهجيات لتطوير المراقبة والرصد على الإنترنت لشبكات وسائل الإعلام الاجتماعية. وهي تقدم مقدمة للمفاهيم ذات الصلة التي ينبغي لنا أن نفهم بها عند دراسة هذه القضايا، فضلاً عن مراجعة لكيفية بناء صورة كاملة للسياق الاجتماعي الفني في دولة أو منطقة، بما في ذلك تواجد الأحزاب المحلية على شبكة الإنترنت، ووسائل الإعلام الاجتماعية ومعدلات اختراق الإنترنت، ووسائل الإعلام المحلية، الانقسامات العرقية والدينية ومجموعة من العوامل الأخرى التي تظهر في حيز الإنترنت.

معلومات حول محتويات ميزة هذا المورد حول المواضيع الرئيسية التالية:

- **التعاون:** يتعين على الباحثين أن يفكروا في الشركاء المحتملين وسبل اختيارهم، سواء من حيث الأنواع المختلفة من المنظمات المحلية، أو المنظمات غير الحكومية الدولية التي تتمتع بالخبرة في هذا المجال، أو الشركات الخاصة التي تتراوح بين أمثلة أصغر حجماً ذات خبرة في هذا المجال إلى شركات ضخمة متعددة الجنسيات تسيطر على منصات وسائل الإعلام الاجتماعية وغير ذلك من التكنولوجيات التي تصل إلى الشبكات العالمية. وسيستعرض قسم من هذه الخيارت المخصص والاعتبارات المحتملة لهذه العمليات التعاونية، مع ذكر أمثلة، مع الإشارة إلى فوائد ومخاطر العمل مع مجموعات مختلفة.

- **المنهجية:** وفيها، يبحث الباحث أساليب مختلفة لجمع البيانات، بما في ذلك اعتبارات تتعلق بمنصات مختلفة،أساليب العمل مع واجهات برمجة التطبيقات لكل منها، واستخراج المحتوى بطرق مختلفة .

- **التخطيط والتهيئة:** يتضمن هذا القسم إرشادات حول تطوير خرائط الشبكات وقراءتها ويقدم الباحث مصطلحات فنية رئيسية فضلا عن طرق بناء خرائط للحيز على الشبكة؛ وأمثلة علي البحوث الميدانية والقيود المحتملة للخرائط نفسها.

- التحليل: يتناول هذا القسم تحليل الأنواع المختلفة من الكيانات والأفراد الذين يهيئون المحادثة يتضمن التوجيه حول هذا الموضوع نظرة عامة على الحسابات الفردية والمؤثرين والمجموعات كما تدرس أدوارها في النظام البيئي عبر الإنترنت، كما تبحث الطرق التي تصبح بها المنظمات الإخبارية وغيرها من الموارد الخارجية مصادر مهمة للمحتوى.

- **المحتوى:**يتناول الباحث المحتوى عبر الإنترنت نفسه، ويبحث في الجوانب المختلفة للمنشورات والتويدات وغيرها من أشكال الوسائط الاجتماعية يصف هذا القسم كيفية اكتشاف الأنواع المختلفة للدعاية الحسابية المرتبطة بشبكة، والتي تتراوح بين شبكات بوت نت ومزارع ترول إلى أشكال أخرى محتملة من التلاعب وهي تأخذ في الاعتبار الأشكال الضارة المختلفة للمحتوى اليدويات التي تتراوح بين المعلومات المضللة والخطاب الذي يحض على الكراهية والأهداف المحتملة.

- **الأدوات:** لدعم هذه التقنيات البحثية، والكتالوجات الإرشادية، وتفحص الأنواع المختلفة من الأدوات المفيدة في تطوير التحليل لمختلف جوانب مراقبة وسائل الإعلام الاجتماعية. وتتضمن هذه القائمة أدوات التجميع على أنظمة أساسية مختلفة، فضلاً عن موارد لتحليل الشبكة، وإظهار البيانات، والبحث في المعلومات ذات المصدر المفتوح.

- **الردود:** في النهاية، يشتمل الدليل على استعراض الاستجابات المحتملة التي يمكن الاطلاع عليها وتعزيزها من خلال هذه التحليلات ويشمل ذلك جمع البيانات للبحث وتطوير الوثائق وإنشاء آليات للإبلاغ مع المنصات والهيئات التنظيمية الحكومية ومنظمات مراقبة الانتخابات. ويختتم الباحث بتوصيات لبناء البحوث في المجالات المقبلة ووضع تقييمات حاسمة للمناطق المتطورة في مجال البحث.

المعلومات العامة

يجب على المحللين عند دخولهم بيئة جديدة أن ينظروا الي العديد من الجهات الفاعلة السياسية والاجتماعية والشبكات والمجموعات ، فضلاً عن الأنظمة التقنية. ويتعين عليهم أن يدرسوا جوانب مختلفة من البيئات الإعلامية والاجتماعية والسياسية التي تعمل فيها، مع مراعاة شبكات وتنظيم البلاد ذاتها، والمنطقة الأوسع، ومكانتها في النظام العالمي ولا يتم تمرير المعلومات دوماً عبر شبكات الاتصال على شبكة الإنترنت أو الشبكات الإعلامية التقليدية؛ بل إن الكثير منها موجود بكلام الناس، من خلال الشائعات، ووسائل الإعلام التقليدية، وغير ذلك من الطرق المختلفة , ولكن هذه المناقشات كثيراً ما تتخلل على شبكة الإنترنت، حيث يمكن تتبعها وفهمها بشكل أفضل الآن.

وتستخدم مجموعات مختلفة على المستوى الدولي أساليب للتلاعب بالتصور العام للمرشحين والقضايا، وإضعاف الثقة في العمليات الديمقراطية، وإرباك الناخبين بشأن مواقع الاقتراع، أو تسجيلهم، أو النظام الانتخابي ذاته . ومن بين الأغراض المركزية لهذا النوع من البحوث المساعدة في تسليط الضوء على الكيفية التي تعمل بها شبكات الدعاية الحسابية[1] على شبكة الإنترنت، والمساعدة في الكشف عن كيفية عملها، وتوثيق حالات البحث، وربما تنبيه السلطات أو شركات وسائل الإعلام الاجتماعية إلى التلاعب والاستغلال على شبكة الإنترنت .إن الكشف عن التشغيل الآلي، والحسابات الزائفة، والمحتوى المزيّف، والمصادر الرديئة للمعلومات، وغير ذلك من أشكال التلاعب، قد يكون جميعها أهدافاً من هذا النوع من المشاريع لوصف الدعاية الحسابية في حالات مختلفة.

يتعين على محللي وسائل الإعلام الاجتماعية الذين يتطلعون إلى قضايا الإعلام في مختلف أنحاء العالم أن يضعوا في حسبانهم العديد من العوامل عند تقديم التقارير. كما يتعين عليها أن تنشر الأدوات, وأن تبحث في القوانين والمؤسسات التي تحكم الحير المعلوماتي, وأن تستخدم أساليب غير متصلة بالإنترنت مثل المقابلات مع المسؤولين الحكوميين، والأحزاب، ووسائل الإعلام، والمرشحين لرسم الصورة الأكثر دقة للتضليل في بيئة الانتخابات.

إن التضليل من المعلومات يشكل موضوعاً صعباً للالتقاط لأنه في عموم الأمر ليس مصطلحاً جيد التحديد ومن العناصر الرئيسية للتضليل هو مفهوم القصد، حيث أن التضليل يتم بالغرض منه الخداع، في حين أن التضليل هو محتوى غير صحيح دون أن يكون القصد من تزييف المعلومات ضروريا. للمزيد من المعلومات عن التعريفات ومعلومات أساسية أخرى عن هذه المواضيع، راجع تقرير "بيانات والمجتمع" عن الأكاذيب واضطراب "مشروع المعلومات" الأول اضافة الي دعم المعهد الوطني الديمقراطي للبحث حول نزاهة المعلومات والخطاب السياسي المدني[2]، الذي تمت ترجمته باللغة الألبانية والعربية والإنجليزية والفرنسية والروسية والصربية والأسبانية وترد في الفرع المراجع مصادر أخرى. وبالإضافة إلى المعلومات المضللة، يتعين على الباحث أن يفكر في العديد من أنواع المحتوى المختلفة، بعضها حميدة، فضلاً عن خبيثة أو إيجابية.

ومن المهم أيضاً النظر في دور وسائل الإعلام، وكيفية عملها داخل أنظمة وسائل الإعلام على الإنترنت ووسائل الإعلام الاجتماعية والتأثير عليها وبالنظر إلى المصادر الرئيسية للصحف المطبوعة والإذاعية والتلفزيونية، وقم بتقييم حصتهم في السوق، والروابط مع الأحزاب السياسية، ومشاركة القطاعين العام والخاص، وكيف تؤثر على فضاء الإنترنت. العديد منهم يتمتعون بقدر كبير من التواجد على شبكة الإنترنت، وخاصة عندما يرتبط الأمر بحزب سياسي كبير، وقد يكون لهم تأثير كبير على خط التحرير والتفضيلات السياسية لأي منظمة, إضافة إلى الرغبة في اكتساب دعم فيروسي، إما عضوي أو متولد بشكل مصطنع . ويمكن اعتبار الاستقطاب في هذا السياق حاسما فيما يتعلق بالشبكات ، وقد يصبح من الواضح بسهولة عندما تتشكل جماعات أكبر ومتميزة في المعارضة والتحالف، وهو ما قد يشير أيضاً إلى مستويات الدعاية الحسابية والمعلومات المضللة التي تظهر.

كثيراً ما يتضمن خطاب الكراهية معلومات مضللة للتلطيخ بأهداف الحملات بهجمات زائفة. ويستهدف هذا خاصة النساء والأقليات والفئات المستضعفة الأخرى. يلاحظ فريق المرأة والنوع الاجتماعي و البحث حول العنف ضد المرأة في السياسة والديمقراطية التابع للمعهد الديمقراطي الوطني للشؤون الدولية أن "عندما يتم توجيه الهجمات ضد النساء الناشطات سياسياً عبر الإنترنت، يضخم التوسع في منصات وسائل التواصل الاجتماعي آثار الاعتداء النفسي من بجعلها تبدو مجهولة المصدر ولا حدود لها ومستدامة، وهو ما من شأنه أن يقوض شعور النساء بالأمن الشخصي بطرق لا يجرها الرجال. يتم تعبئة العديد من الجهات الحكومية وغير الحكومية التي ترتكب العنف ضد المرأة عبر الإنترنت عبر الشبكات عبر الوطنية. لقد أصبح سوء استخدام – من قبل الدول والمنظمات والأفراد – للحريات نفسها التي يفترض أن تتيحها فضاء المعلومات ، أحد أكبر التهديدات لسلامتها ". (Zeiter et al.، 2019، 4) ونتيجة لذلك ، يجب أن يكون الباحثون على دراية بكيفية تعرض النساء بشكل خاص لهذه الهجمات عبر الإنترنت ، والنظر في طرق توثيق وإبلاغ هذا النوع من الانتهاكات إلى المنصات. وقد جرب المعهد الوطني استخدام مصطلحات خطاب الكراهية لدراسة الشبكات الاجتماعية مع دراسات الحالات الإفرادية في كولومبيا وإندونيسيا وكينيا. إن أساليب تطوير هذه الرموز ودراسات الحالة المفصلة في التقرير الخاص بهذا البحث "تغريدة : تحليل العنف ضد المرأة في السياسة".

1 ووفقا لمعهد أوكسفورد للإنترنت، الذي ساعدت مجموعته البحثية في تحديد المصطلح وتحريك القسم الأعظم من الأبحاث التي تدور حول هذه القضايا: "الدعاية الحاسوبية هي إستخدام الخوارزميات، والأتمتة، والسخرية البشرية لتوزيع المعلومات المضللة عمدا على شبكات وسائط التواصل الاجتماعي" (وواللي وهوارد 2017 ، 4).

2 انظر الصفحة https://www.ndi.org/publications/supporting-information-integrity-and-civil-political-discourse

عند وضع استراتيجيات لجمع البيانات ضع في اعتبارك كيفية عمل هذه المعلومات باستخدام خطاب الكراهية والدعاية الحسابية في السياق الذي تعمل فيه . ستساعد المحتويات التالية لهذا البحث المحللين على التعرف على كيفية التعرف على هذه الحملات والشبكات والمستخدمين على الإنترنت ، ولكن يتعين على الباحثين أن يفهموا على نطاق واسع أي نوع من الدعاية الحسابية، أو الخطاب الضار، أو غير ذلك من الأنماط التي يبحثون عنها قبل المتابعة .

ومن المهم أيضاً أن يكون هناك فهم واسع النطاق للبيئة التنظيمية المحلية (بما في ذلك تنظيم تمويل الانتخابات والحملات الانتخابية، والذي يمكن أن يساعد في إعداد التقارير) وقواعد اعتدال المحتوى والسياسات التي تحكم البرامج التي تتم دراستها. وهذا يعني أن الباحثين يستطيعون تصميم دراساتهم بشكل قانوني وأخلاقي، وتنبيه الشركات – والحكومات المحتملة – إلى إساءة المعاملة وغيرها من الإجراءات غير القانونية والسلبية على الإنترنت. وهذا من شأنه أيضاً أن يساعد الباحثين على فهم البلدان التي يدرسون فيها بشكل أفضل.

ويمثل فهم شروط الخدمة لمختلف شركات وسائل الإعلام الاجتماعي جانباً بالغ الأهمية من جوانب هذا التقرير فيما يلي المعايير الرئيسية التي يجب فهمها بالنسبة إلى Facebook وTwitter وYouTube وترد أدناه روابط إلى سياساتهم المختلفة.

معايير منصة Facebook	• معلومات موضوعية حول الإبلاغ عن سيناريوهات مختلفة, هنا. • "الإبلاغ عن صفحة ملصمة عن شخصية عامة" هنا. • "كيفية الإبلاغ عن الأشياء" مقسمة حسب نوع مختلف من المنشورات، هنا. • "كيف يمكنني وضع علامة على منشور كأخبار خاطئة؟" هنا.
قواعد منصة Twitter	• نظرة عامة على انتهاكات التقارير, هنا. • كيفية الإبلاغ عن تغريدة أو حساب مسيء أو رسالة فردية هنا. • الإبلاغ عن حساب انتحال شخصية, هنا. • الإبلاغ عن تعليمات البريد الإلكتروني العشوائي, هنا.
توجيهات منصة YouTube	• كيفية الإبلاغ عن محتوى غير ملائم، مقسماً حسب نوع المنشور، هنا. • "إبلاغ عن توقع بحث يوتيوب", هنا. • "خيارات إبلاغ أخرى," هنا. • "كيفية الإبلاغ عن محتوى غير هام أو مخادع"، هنا (نهاية الصفحة). • أداة إعداد تقاريراليوتيوب, هنا.

تطلب بعض الشركات، مثل Facebook، من المستخدمين التعريف بأنفسهم مستخدمين معلوماتهم الحقيقية، لذا فإن تحديد حساب ليس شخصًا حقيقيًا أو مجموعة مرتبطة بهذا الحساب الزائف يمكن أن يؤدي إلى الإزالة. بينما المنصات الأخرى، مثال Twitter، لا تحظر إخفاء الهوية ، ولكن لديهم حظرًا ضد خطاب الكراهية أو التضخيم الاصطناعي الذي يمكن تحديده والإبلاغ عنه من خلال البحث. يجدر التعرف على الرموز والآليات المختلفة للإبلاغ عن المحتوى ، والموضحة في الجدول أعلاه.

يعد إعداد التقارير أمرًا مهمًا ، مثل توثيق الحملات والحسابات والمحتوى ذي الصلة حتى يمكن التحقق من التقارير. ضع في اعتبارك استخدام أنظمة يمكن نسخها احتياطيًا واستشارتها بسهولة ، من حيث التعليقات التوضيحية وسير العمل. كما ذكرنا ، فإن معايير مجتمع Facebook يتوفر آليات للإبلاغ عن الحسابات الخاطئة والأشكال السلبية للمحتوى ، ولطلب التحقق من المحتوى الضار وربما التخلص منه.

تسمح قواعد Twitterبأشكال مختلفة للإبلاغ عن انتحال الهوية والرسائل غير المرغوب فيها وأشكال معينة من الكلام الذي يحض على الكراهية أو غير ذلك. تركز سياسات YouTube على الوسائط وحقوق الطبع والنشر ، بالإضافة إلى أشكال التعبير الصريحة أو البغيضة أو غيرها من المحتويات الضارة. لا تمنحك اختصاصات جوجل فكرة عن كيفية عدم الإبلاغ عن الحسابات فحسب ، بل التحقيق أيضًا في الحسابات المرتبطة بخدمة البث المباشر بصفتك صاحب حساب يوتيوب.

وفيما يتعلق باللوائح الحكومية، ينبغي على المحللين أن ينظروا في قوانين حماية البيانات مثل اللائحة العامة لحماية البيانات في الاتحاد الأوروبي، والتي تتضمن جوانب تغطي جمع المعلومات الشخصية في أوروبا أو عن الأوروبيين في أي مكان، فضلاً عن الشركات التي تعمل هناك[3] .ويمكن لجمع البيانات من مجموعات خاصة ومن المستعملين والشبكات أن يلغي هذه القوانين وكذلك شروط خدمة المنصات.

يمكن أن تساعد شبكات مثل Design 4 Democracy Coalition (D4D Coalition) في الدفاع عن المبادئ الديمقراطية في شركات التكنولوجيا – على سبيل المثال ، لجذب انتباه المنصات على الحملات ذات التأثير الأكبر للانتخابات أو خطاب الكراهية أو أغراض أخرى إلى انتباه المنصات.

يتألف تحالف (D4D Coalition): من منظمات المجتمع المدني الدولية والمحلية التي تشارك في شركات التكنولوجيا لدمج ودعم المبادئ الديمقراطية، بما في ذلك سياق الإشراف على المحتوى، ووضع السياسات، واعتبارات المنتجات. التحالف (بقيادة المعهد الديمقراطي الوطني للشؤون الدولية، والمعهد الجمهوري الوطني , والمؤسسة الدولية للأنظمة الانتخابية والمعهد الدولي للديمقراطية ومساعدات الانتخابات) تربط بين المجتمع المدني وأصحاب المصلحة في مجال الديمقراطية في سياقات متنوعة والعديد من شركات التكنولوجيا الأكثر نفوذاً (بما في ذلك Facebook وMicrosoft وTwitter) لتشجيع تبادل المعلومات والنهوض بالاستراتيجيات لتعزيز سلامة المعلومات وحماية العمليات الديمقراطية. ويمكن دعم وتعزيز الجهود المبذولة من هذا القبيل من خلال توثيق وتحليل وإعداد تقارير متينة باستخدام الأدوات والأساليب والتكتيكات الموضحة هنا.

ولمزيد من التفاصيل حول وضع استراتيجيات لتحليل البيانات المتعلقة بوسائل التواصل الاجتماعي الخاصة بالانتخابات, يمكنك مراجعة مستند توجيهات المعهد الديمقراطي الوطني للشؤون الدولية كوثيقة إرشادية حول التضليل والنزاهة الانتخابية[4]، بالإضافة إلى دليل دعم الديمقراطية للمجتمع المدني حول مراقبة وسائل التواصل الاجتماعي للانتخابات[5]. تغطي هذه الأدلة المنهجيات والاعتبارات التنظيمية للمراقبين من حيث مراقبة وسائل التواصل الاجتماعي وإمكانية دمج البيانات التي يتم جمعها عبر الإنترنت في بعثات مراقبة الانتخابات التقليدية.

العمل على جمع البيانات

إن جمع البيانات يشكل الخطوة الأولى نحو تحليل دقيق للنشاط على شبكة الإنترنت حول الانتخابات. وكمحلل على أرض الواقع، فإن الخطوة الأولى لجمع البيانات هي دراسة مسح على شبكة الإنترنت/مواقع التواصل الاجتماعي في المنطقة التي تراقبها. وتشمل المسائل الهامة التي ينبغي النظر فيها ما يلي :

- **ما هي المنصات الأكثر شعبية في المنطقة؟ ما هي معدلات اختراق المنصة المختلفة؟** إن منصات إحصائيات عالم الإنترنت، أو الاتحاد الدولي للاتصالات، وفريدوم هاوس عن تقرير الإنترنت، أو Facebook ذاته، من الممكن أن تشكل موارد جيدة هنا. فالدولة التي تتمتع بمعدلات مشاركة مرتفعة على Facebook ولكن معدلات اختراق منخفضة على Twitter(مثل تايوان) من المرجح أن تسفر عن أعظم الرؤى القيمة على Facebook .
- **ما هي مواقع ويب التي تحظى بشعبية كبيرة بالنسبة للأخبار؟ أي مما يلي من الوسائط القديمة؟ أي من هذه الإبداعات أحدث؟**
- **أية هاشتاغات – وسمات - أكثر صلة بالانتخابات المعنية؟ ما هي الحسابات الرسمية التي تمثل الأحزاب والمرشحين وحملاتهم في هذه الانتخابات؟** وفي كثير من الأحيان يكون للمرشح أكثر من حساب أو صفحة ذات صلة بالانتخابات مثل حساب تويتر شخصي وحساب تويتر الرسمي للحملة الانتخابية. ويعد إعداد قائمة بهذه البيانات خطوة أولى جيدة لجمع البيانات ذات الصلة.

بعد تمييز أهم منصات ومواقع التواصل الاجتماعي في البلاد، تكون مستعدا للبدء بجمع البيانات على المنصات والمواقع ذات الصلة. في هذا القسم، سنبحث في الخيارات المتاحة للباحثين لجمع البيانات.

3 https://gdpr.eu/

4 https://www.ndi.org/publications/disinformation-and-electoral-integrity-guidance-document-ndi-elections-programs

5 يتم دعم الديمقراطية من خلال اتحاد يتألف من SOFRECO و المنظمة الدولية للتقارير عن الديمقراطية و المعهد الديمقراطي الوطني للشؤون الدولية https://democracy-reporting.org/wp-content/uploads/2019/10/social-media-DEF.pdf

أساليب جمع البيانات

عند جمع بيانات وسائل التواصل الاجتماعي عبر الإنترنت، تتوفر للباحثين خيارات متعددة . الخيارات الثلاثة الرئيسية هي استخدام أدوات تجميع تابعة لجهة خارجية، أو التفاعل مباشرة مع واجهة برمجة التطبيقات الخاصة بمنصة ما، أو المشاركة في "كشط الويب " سنستكشف التفاصيل والاختلافات بين هذه الأساليب الثلاثة في هذا القسم.

أدوات خارجية (أدوات غير مباشرة)

بالنسبة للباحثين الذين يستخدمون جدولاً زمنيًا ضيقا أو الذين لا يملكون القدرة على التفاعل مع مواقع الويب من خلال رمز الكمبيوتر ، يمكن أن تكون الأدوات الخارجية خيارًا مفيدًا . تتفاعل هذه الأدوات عادة مع واجهات برمجة التطبيقات الخاصة بمنصة هدف واحدة أو أكثر بشكل غير مرئي في الخلفية، وتقدم البيانات في تخطيط رسومي سهل الاستخدام مثل لوحة المعلومات . بالنسبة لمنصة Facebook، الذي لا يسمح حاليًا للباحثين الخارجيين أو الشركات باستخدام واجهة برمجة التطبيقات الخاصة به، فإن ميزة زاوية التجميع (CrowdTangle) هي أحد أفضل الخيارات لمراقبة الوصول إلى الصفحات والمجموعات وعناوين مواقع الانترنت على Facebook، كما يعرض للمستخدمين إمكانية الوصول الي مواقع انترنت علي حسابات Instagram وYouTube و Reddit. ويُتيح (CrowdTangle)[6] للمستخدم رؤية الوقت الفعلي المقدر لعدد التفاعلات في منشور،أو صفحة أو عنوان URL على هذه المنصات، والتي تعتبرطرقا مفيدة لمراقبة المحتوى الرائج عبر منصات متعددة على أساس يومي.

تتطلب بعض الأدوات الخارجية مثل شركة سيسوموس وشركة براندواتش اشتراكًا مكلفًا مدفوع الثمن لاستخدامه، ولكنها توفر إمكانية الوصول إلى كمية كبيرة من البيانات التي قد تكون مفيدة لمراقبة حملات الهاشتاق – الوسمات - والمحتوى الآخر الأكثر رواجًا على وسائل التواصل الاجتماعي.

بالإضافة إلى تصوِّر البيانات من هذه الأدوات على مستعرض ويب، يمكن غالبًا تصديرها إلى ملف قابل للقراءة بواسطة جهاز، مثل ملف قيم تفصلها الفواصل والذي يمكن معالجته بدوره بواسطة عالم بيانات للاستعلام عن البيانات بطرق جديدة ومفيدة.

فيس بوك: سيسوموس /براندواتش

توتير: احصائيات توتير

الوصول المباشر- ما الفرق بين واجهات برمجة التطبيقات و كشط الويب ؟

لمزيد من التفاعل المباشر مع الأنظمة الأساسية ومواقع الويب ، يتوفر للباحثين خيارين متاحين: استخدام واجهة برنامج التطبيق (API) أو جمع المعلومات مباشرة من شفرة مصدر صفحة الويب ، وهي ممارسة تعرف باسم "كشط الويب". من المهم ملاحظة الفرق بين واجهات برمجة التطبيقات وكشط الويب - إن سحب البيانات من واجهات برمجة التطبيقات قانوني وأخلاقي في معظم الحالات حيث يتم تنظيم بيانات واجهة برمجة التطبيقات عمدًا بواسطة منصات ويتم تنظيمها بحيث لا تنتهك حقوق المستخدمين. بينما كشط صفحات الويب، في العديد من الحالات، يشكل انتهاكاً لشروط الخدمة، ومن الصعب تنظيمه. في كثير من الحالات، قد يكون كشط الويب غيرقانوني.

من المهم ملاحظة الفرق بين هاتين الطريقتين لجمع البيانات. من المهم أيضًا معرفة أنك ستسمع من حين لآخر الباحثين أنفسهم يشيرون عن طريق الخطأ إلى البيانات التي تم استردادها من واجهة برمجة التطبيقات باعتبارها "مفكوكة". التمييز مهم - ليس فقط لأسباب عملية مثل توفير الوقت والجهد ، ولكن أيضًا للانتهاكات الأخلاقية والقانونية التي يمكن أن تأتي من كشط الويب.

مع وضع هذا القياس الأساسي للفارق بين النهجين في الاعتبار ، دعونا ندفع في تفاصيل واجهات برمجة التطبيقات وكشط الويب.

واجهات برمجة التطبيقات (API)

بالنسبة للباحثين المهتمين بنهج أكثر عملية لجمع البيانات، فإن العديد من المنصات لديها شكل مباشر أكثر للوصول إلى البيانات في شكل واجهات برمجة التطبيقات وبشكل عام، تمثل واجهات برمجة التطبيقات طريقة للمستخدمين للتفاعل بسهولة مع موقع ويب أو منصة تواصل اجتماعي من خلال رموز الكمبيوتر. وهذا من شأنه أن يمكن المستخدم من معالجة المزيد من البيانات بسرعة، وبالتالي توليد رؤى أعمق للنشاط عبر الإنترنت، مقارنة بما كان ليستطيع القيام به يدويًا.

6 تتوفر شبكة CrowTangel حاليا للأكاديميين والباحثين على أساس انتقائي. يمكنك أنت وفريقك التقدم على موقع CroadTangle.com. وتتضمن النسخة الكاملة البيانات الحالية والتاريخية على فيسبوك وإنستاجرام. كما يتوفر ملحق CroadTangle المجاني على الموقع الشبكي. بأخذ هذا الملحق عنوان URL كمدخل ويقدم أحدث ٥٠٠ تدوينة عامة جذبت إلى أي حد كبير مستندا إلى عنوان URL على فيسبوك وإنستاجرام وريديت وتويتر. يمكن أن يكون كل من المنصة الكاملة و ملحق CrowTangel مفيدا للتحقيقات.

أنواع واجهات برمجة التطبيقات

هناك ميزتان من ميزات واجهة برمجة التطبيقات على وجه الخصوص ذات صلة يجب أن تكون على دراية بها قبل بدء جمع البيانات: الانفتاح والإطار الزمني.

- **الواجهات المفتوحة:** تأتي واجهات برمجة التطبيقات في أشكال مختلفة - تسمح واجهات برمجة التطبيقات المفتوحة بجمع البيانات من قبل أي شخص، بينما تتطلب واجهات برمجة التطبيقات المصادق عليها من المستخدم الخضوع لشكل من التحقق قبل السماح بجمع البيانات.

- **واجهات برمجة التطبيقات المفتوحة:** تطبيق فينمو المستخدم للمدفوعات الإلكترونية في الولايات المتحدة، لديه واجهة برمجة تطبيقات عامة تسمح لأي شخص بعرض عدد من أحدث المعاملات العامة على التطبيق. يمكنك العثور على قوائم الواجهات المفتوحة من خلال الإنترنت - مثل هذه القائمة على جيثب (Github). كما يحتوي Any-api.com على قائمة بعدة واجهات برمجة تطبيقات متاحة للجمهور للمستخدمين المهتمين، والكثير منها مفتوح.

- **واجهات برمجة التطبيقات المصادق عليها:** تتطلب معظم هذه الواجهات التي ستتعامل معها لجمع بيانات التواصل الاجتماعي (ريديت ,وتوتيروغيرها) مصادقة المستخدم قبل أن تتمكن من جمع البيانات.

- **الإطار الزمني:** تقوم مواقع الويب والمنصات أيضًا عادة بتركيب واجهات برمجة التطبيقات الخاصة بها بشكل مختلف وفقًا للوقت.

- **جمع البيانات التاريخية:** تسمح لك معظم واجهات برمجة التطبيقات بجمع بعض أشكال البيانات التاريخية علي مواقعها ريديت /وتوتير يسمح بذلك بشكل ملحوظ يتيح لك نموذج واجهة برمجة التطبيقات هذا ، الذي يطلق عليه واجهة برمجة تطبيقات تاريخية ، إمكانية سحب البيانات التي تم إنشاؤها بأثر رجعي قبل إجراء الاستعلام. الأهم من ذلك أن البيانات التي تم نشرها بعد إجراء الاستعلام غير متاحة للتجميع.

- **تدفق البيانات الحالية:** يُشار إلى نقل البيانات وتنزيلها عند حدوثها في الوقت الفعلي باسم عملية الدفق. . إذا كانت منصة تويتر أساسية ذات صلة في الانتخابات أو الفترة التي تخطط لمراقبتها ، فمن المرجح أن يكون البث هو أفضل خيار لك لجمع البيانات. عند دفق البيانات ، تقوم بجمع التغريدات في الوقت الحقيقي وفقًا لاستعلام محدد (على سبيل المثال ، جميع التغريدات التي تستخدم علامة الوسم # ، أو جميع التغريدات التي تشير إلى حسابات الفائدة ، مستشهدة بعناوين URL ذات الاهتمام ، وما إلى ذلك)).

حدود واجهة برمجة التطبيقات

حرصاً على الحفاظ على خصوصية المستخدم وسلامته، تعمل معظم منصات وسائل التواصل الاجتماعي على الحد من كمية البيانات المسموح لأي مستخدم بتجميعها. نستكشف بعضًا من هذه القيود هنا لتعريفك بالمشكلات التي قد تواجهها عند جمع البيانات.

- **قيود الحجم:** تحد معظم واجهات برمجة التطبيقات من حجم البيانات التي يستطيع المستخدم جمعها على سبيل المثال، عندما تقوم ببث البيانات في الوقت الحقيقي على تويتر، فإن النظام الأساسي يغطي كمية البيانات التي يستطيع المستخدم جمعها إلى 1% من بيانات البث العالمية.

- **قيود المعدل:** تعتبر أكثر أنواع القيود شيوعاً التي ستواجهها عند سحب بيانات واجهة برمجة التطبيقات. تمتلك معظم واجهات برمجة التطبيقات حدودًا للمعدل للتأكد من أن مستخدمًا واحدًا أو تطبيقًا واحدًا لا يمكنه تنزيل كمية زائدة من البيانات (كما هو محدد بواسطة النظام الأساسي أو موقع الويب المعين). علي سبيل المثال يحد تويتر من عدد التغريدات التي يمكن للمستخدم الواحد تنزيلها ضمن أي نافذة لمدة 15 دقيقة.

- **إزالة قيود البيانات:** كما تم التطرق إليه بإيجاز أعلاه، تميل الأنظمة الأساسية إلى إزالة المحتوى الذي ينتهك القواعد والأنظمة المحددة للاستخدام - حيث تتضمن هذه الأنظمة أسماء مختلفة (معايير المجتمع وشروط الخدمة، وما إلى ذلك). وهذه الملاحظة مهمة بشكل خاص في السياقات الانتخابية التي من المرجح أن يحدث فيها سلوك مشين، وخاصة أن البيانات المتدفقة على تويتر تمكنك من جمع وحفظ البيانات عن الجهات الفاعلة الشائنة في الوقت الحقيقي والتي يمكن إزالتها من النظام الاساسي في وقت لاحق. وبمجرد إزالة البيانات عن هذه الجهات الفاعلة تصبح غير متاحة . إذا كان فريقك يريد التقاط الممثلين السيئين والمعلومات المضللة والمحتويات الأخرى لتحليلها لاحقًا ، فإن البث في الوقت الفعلي يزيد من فرصك في القيام بذلك.

- **التقيد بنوع البيانات :** وبالمثل ، ستحد معظم الأنظمة الأساسية من نوع البيانات التي يمكن للمستخدم جمعها من منصته. ستسمح لك واجهة برمجة تطبيقات Twitter بجمع معلومات معينة عن المستخدم المستهدف (مثل عدد منشوراته وعدد المتابعين وتاريخ إنشاء الحساب) ، ولكنها لن تسمح لك بالوصول إلى أنواع أخرى من البيانات المقيدة (مثل معظم المستخدمين عنوان IP المستخدم بشكل متكرر). لا يسمح Facebook حاليًا للباحثين بجمع معلومات حول المستخدمين أو الصفحات الأخرى من خلال واجهة برمجة التطبيقات الخاصة به على الإطلاق. من المفيد أن تكون على دراية بنوع البيانات التي يمكن جمعها على الأنظمة الأساسية المستهدفة عند تصميم مشروع بحثي.

- **التقيد بوقت البيانات :** يقصر Twitter المستخدمين على البيانات التي حدثت في الأيام 9-7 الماضية عند جمع البيانات التاريخية من خلال واجهة برمجة تطبيقات بحث Twitter. يمكن شراء البيانات الأقدم من 9-7 أيام على Twitter من موفري البيانات مثل GNIP ، ولكن لا يمكن جمعها من خلال الوصول القياسي لواجهة برمجة التطبيقات.

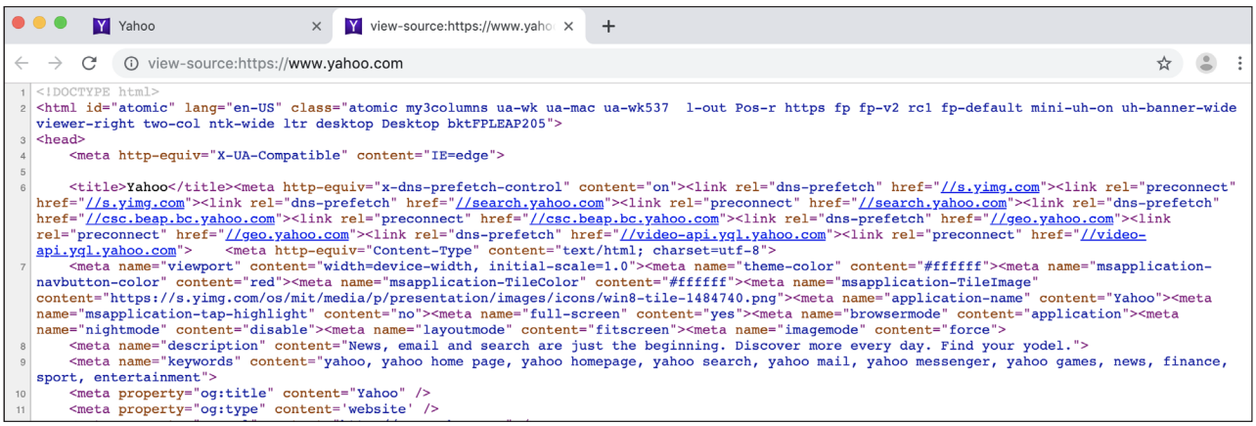
من الجوانب المثيرة لواجهات برمجة التطبيقات أنها ليست خاصة بلغة معينة. يمكن سحب بيانات واجهة برمجة التطبيقات باستخدام Java أو Python أو R أو Ruby أو Perl أو أي لغة برمجة أخرى قد تفضلها أنت أو فريقك الفني. في حين أن هذا هو الحال ، هناك حزم محددة بلغات البرمجة الشائعة التي تبسط عملية الاستعلام عن من خلال معالجة بعض التعقيدات بالنسبة لك.

يتوفر رمز مثال مع توضيحات توضح كيفية استخدام حزمة التغريد لجمع البيانات من واجهة برمجة تطبيقات البحث و البث على Twitter في نهاية هذا الدليل الميداني في الملحق: مثال رمز واجهة برجمة التطبيقات - جمع البيانات من واجهات برمجة تطبيقات البحث و البث مع حزمة التغريد.

كشط الويب

بينما تقدم واجهات برمجة التطبيقات طريقة مبسطة لجمع البيانات من نظام أساسي أو خدمة ، فهي ليست الخيار الوحيد لجمع البيانات. كشط الويب هو عملية استخراج شفرة المصدر من موقع ويب مستهدف واستخراج البيانات ذات الصلة. كل صفحة على الإنترنت هي نتيجة شفرة المصدر التي تتكون منها HTML - ولغات الويب الديناميكية الأخرى ، والنصوص البرمجية ، والارتباطات التشعبية ومصادر الوسائط. تسمى عملية متصفح يأخذ نصًا برمجيًا ويحوله إلى صفحة ويب مرئية وتفاعلية عرض. يمكنك عرض شفرة المصدر الأساسية لأي صفحة ويب في معظم المتصفحات الحديثة - كل من Google Chrome و Mozilla Firefox و Safari و Brave و Opera جميعها تحتوي على هذه الميزة.

على سبيل المثال ، عند استخدام Google Chrome ، يمكنك النقر بزر الماوس الأيمن على أي صفحة ويب تم تحميلها (أو "تقديمها") في متصفحك والنقر على خيار "عرض مصدر الصفحة". سيفتح Chrome علامة تبويب جديدة تعرض لك شفرة HTML و CSS المستخدمة لتحميل صفحة الويب التي تعرضها. فيما يلي مثال على ذلك من موقع yahoo.com.



لقطة شاشة لكود مصدر HTML و CSS الأساسي لـ yahoo.com (لقطة تم التقاطها في أوائل أغسطس 2019). يتيح Google Chrome والمتصفحات الحديثة الأخرى للمستخدمين عرض شفرة المصدر لأي موقع ويب يزورونه - رمز المصدر هذا هو ما يتم استرداده عند "إلغاء" موقع الويب.

مخاوف أخلاقية عند كشط الويب

يمكن أن تكون عملية كشط الويب أداة مفيدة لتحليل صفحات الويب. تجدر الإشارة مع ذلك إلى أن كشط الويب يعد انتهاكًا لبنود الخدمة الخاصة بمنصات وسائل التواصل الاجتماعي ، ويمكن أن يكون من السهل خرق القانون عند إلغاء صفحات الويب. لهذا السبب ، من المهم دائمًا مراعاة خصوصية المستخدم والشروط والأحكام والقوانين ذات الصلة عند كشط مواقع الويب. من المهم دائمًا التأكد من أن عملية جمع البيانات الخاصة بك أخلاقية وقانونية. **يجمع العديد من الباحثين البيانات فقط من خلال واجهات برمجة التطبيقات ويختارون عدم مسح المواقع الإلكترونية لهذه الأسباب.**

عند استخدام البيانات التي تم جمعها من قبل فرق أو أدوات أخرى - وهو سيناريو متكرر للعديد من الفرق - من المهم أيضًا التأكد من الحصول على هذه البيانات بطريقة أخلاقية. على سبيل المثال ، إذا تم الحصول على البيانات بشكل غير قانوني من خلال الكشط أو القرصنة ، فمن غير المستحسن استخدامها لمشروع بحثي لأسباب قانونية وأخلاقية وسياسية ، وهذا النوع من المشروع يمكن أن يؤدي إلى تداعيات من قبل السلطات الحكومية والشركات. هذه الاعتبارات مهمة يجب مراعاتها قبل البدء في تحليل البيانات.

فروقات المنصة: جدول تلخيصي

	واجهة برمجة التطبيقات التاريخية المتاحة	واجهة برمجة التطبيقات المتدفقة المتاحة	الأدوات الخارجية لجمع البيانات
Twitter	نعم	نعم	CrowdTangle (ملحق)
Facebook	لا	لا	CrowdTangle
Gab	Yes, through Pushshift.io	لا	لا
Instagram	لا	لا	CrowdTangle
Reddit	نعم	نعم (حزم معينة تحوي هذه الخاصية مثل Reddit SSE)	CrowdTangle (ملحق) Pushshift.io
YouTube	نعم	لا	
Telegram	نعم	لا	Telethon Pushshift.io
VKontakte	نعم	لا	
WhatsApp	نعم تتيح واجهة برمجة تطبيقات WhatsApp business الاتصالات الآلية من الشركات للعملاء. لا يتم استخدامه عادةً لنوع التحليل الذي نناقشه في هذا الدليل.	لا	تتيح العديد من أدوات الجهات الخارجية التحليل الإحصائي أو التصور لدردشات WhatsApp.[7] • ChatAnalyzer • محلل WhatsApp Chat • Chatilyzer • WhatsAnalyzer

أدوات مفيدة لجمع البيانات في سياق نزاهة الانتخابات:

- **حزم واجهة برمجة تطبيقات Twitter**
  - حزم R: twitteR, rtweet
  - حزم Python: tweepy, python-twitter
- **CrowdTangle:** تطبيق CrowdTangle و CrowdTangle Extension هو أفضل الأدوات المستخدمة في تحليل بيانات facebook و Instagram .
- Pushshift.io هوموقع يقوم بأرشفة البيانات من منصات وسائل التواصل الاجتماعي. Reddit, Gab Twitter and Telegram. حصل مؤسس ومشغل شركة Pushshift , جيسون باوم غارتنر, على بياناته من خلال الوصول إلى واجهة برمجة التطبيقات, مما يجعل استخدم البانات أمناً من منظور أخلاقي .
- **أداة البث reddit:** https://github.com/pushshift/reddit_sse_stream

7 من المهم أن تكون واضحًا بشأن آثار الخصوصية على استخدام هذه الأدوات. على وجه الخصوص ، تريد أنت وفريقك التأكد من عدم تعرض الدردشات الخاصة أو حفظها من قبل طرف ثالث عندما تواجه أدوات تحليل WhatsApp.

- **سحابة وسائل اعلام معهد ماساتشوستس للتكنولوجيا:** هي أداة لتجميع الأخبار يمكن أن تكون مفيدة لاستكشاف تغطية المواضيع التي تهم في منافذ متنوعة. يصف الموقع ، الذي يصف الأداة ، "نجمع البيانات من أكثر من 50000 مصدر إخباري من جميع أنحاء العالم وبأكثر من 20 لغة بما في ذلك الإسبانية والفرنسية والهندية والصينية واليابانية. تساعد أدواتنا في تحليل وتقديم وتصور المعلومات حول محادثات الوسائط على ثلاثة مستويات أساسية: الانتباه والتغطية القصوى للقضايا وتحليل الشبكات واستخدام اللغة المجمعة. "

تحليل البيانات والشبكات

بناء تمثيلات مرئية للشبكات الاجتماعية وتحليل العلاقات داخلها هي عملية تعرف باسم تحليل الشبكات الاجتماعية (SNA). يُشار إلى هذه العملية أيضًا باسم بناء خريطة شبكة اجتماعية ، أو "رسم خرائط" شبكة اجتماعية. على الرغم من أنه ليس من الضروري دائمًا استخدام تحليل الشبكات الاجتماعية لفهم مجال الإعلام عبر الإنترنت حول الانتخابات ، إلا أنها يمكن أن تكون طريقة مفيدة لتكوين رؤى مفيدة حول التأثير داخل مساحة معينة من مجتمع وسائل التواصل الاجتماعي ، ويمكن أن تكون طريقة مفيدة لتصور هذا المجتمع.

في جوهره ، يعد رسم خريطة الشبكة الاجتماعية عملية تتكون من 5 خطوات:

1. جمع البيانات
2. تحديد العلاقة بالخريطة
3. تشذيب البيانات
4. إنشاء الخرائط
5. تحليل الخرائط

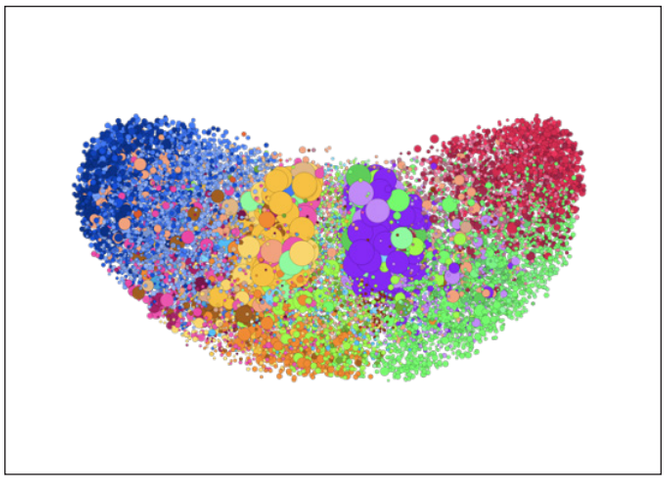
سنناقش كل خطوة من هذه الخطوات بمزيد من التفصيل أدناه.

المصطلحات الأساسية

عند مناقشة وتحليل خرائط الشبكات الاجتماعية ، هناك بعض المصطلحات تساعد على فهم المفردات والممارسات النموذجية لرصد وسائل التواصل الاجتماعي. الشبكات مفيدة للغاية لأنها يمكن أن تمثل العديد من العلاقات في العديد من السياقات المختلفة. في حين أن خرائط الشبكات الاجتماعية ربما تكون الأكثر شيوعًا ، يمكن استخدام الشبكات لنمذجة انتشار الأمراض والفيروسات ، أو رسم خريطة للنشاط العصبي في الدماغ ، أو لتمثيل طرق السفر المحتملة من مدينة إلى أخرى. في حين أن قابلية تطبيق الشبكات على العديد من المجالات المختلفة مفيد ، فإنه ينتج أيضًا حاجة للغة مجردة للتحدث عن الشبكات بغض النظر عن مساحة التطبيق.

المصطلحات الأكثر استخداما هي الرسم البياني و العقدة والحافة.

- **الرسم البياني -** هو مصطلح علوم الكمبيوتر لشبكة تتكون من العقد والحواف. من المهم معرفة هذه الكلمة فقد تصادفها في الأدوات التي تصنع خرائط الشبكة ، أو في مناقشات تلك الأدوات. يمكنك اعتبار كلمة الرسم البياني مرادفة تقريبًا للشبكة.
- **العقدة (أو القمة) -** و هي العناصر التي تشكل شبكة. أحد الأشياء الرئيسية التي يجب معرفتها عن العقد هو أن ما تمثله العقدة يختلف اعتمادًا على الخريطة / التصور الذي تنظر إليه. تمثل العقدة في خريطة شبكة لورنس ألكسندر للمواقع المؤيدة للكرملين نطاقًا ، بينما تمثل كل عقدة دائرية في خريطة Graphika لعام 2018 المشهد السياسي الأمريكي على حساب Twitter.



توضح خريطة Graphika للطيف السياسي الأمريكي لعام 2018 على تويتر العقد. كل دائرة في الخريطة هي عقدة أعلاه تمثل حسابًا فرديًا على Twitter. العقد والحواف هي اللبنات الأساسية للشبكات.

- **الحافة (أو القوس) -** الحواف هي الروابط بين العقد في الشبكة ، ويتم تمثيلها بشكل شائع كخط بسيط بين عقدتين. يمكن أن تمثل هذه الاتصالات مجموعة متنوعة من الأشياء. في نموذج لعدوى المرض ، قد تمثل الحواف انتشار الفيروس من مضيف إلى آخر. في رسم بياني للمطارات في الولايات المتحدة ، قد تمثل الحواف بين عقدتين (المطارات) رحلة مباشرة متاحة بين المطارين. يمكن توجيه الحواف أو عدم توجيهها[8] ، ويمكن أن يكون لها قيمة رقمية مرتبطة بها[9] (غالبًا ما يشار إليها باسم الوزن).

- من المحتمل أن تمثل الحواف في الرسوم البيانية لوسائل التواصل الاجتماعي أحد الأشياء القليلة التالية: متابعة العلاقات أو إعادة التغريد أو الإعجابات. ستكون معظم الشبكات التي تراها على Twitter شبكات متابعة[10] - حيث تعرض الحافة أن أحد المستخدمين يتبع آخر - أو شبكات إعادة التغريد.

رسم خريطة الشبكة الاجتماعية: الخطوات المعنية

في هذا القسم ، سنحدد ونفحص الخطوات الخمس التي ينطوي عليها عمل خريطة الشبكة الاجتماعية التي يتم من خلالها إنشاء رؤى.

1. **جمع البيانات -** كما تطرقنا في القسم السابق ، تتضمن هذه الخطوة جمع البيانات ذات الصلة بالانتخابات المحلية من خلال واجهة برمجة تطبيقات التواصل الاجتماعي أو الأدوات الخارجية. بمجرد جمع البيانات ، سيكون لديك تجمع البيانات الأساسي الذي ستحتاج إليه لإنشاء تصور للشبكة الاجتماعية. من المهم أن ندرك أنه سيتم استخدام أجزاء من هذه البيانات فقط في إنشاء الخريطة - يمكن استخدام نفس مجموعة البيانات الأساسية لإنشاء جميع أنواع الخرائط. كيفية اختيار نوع الشبكة / العلاقات التي تهتم بها كثيرًا وما هي أجزاء البيانات الأكثر صلة هي عمل الخطوتين 2 و 3.

2. **تحديد العلاقة بالخريطة -** في خريطة الشبكة الاجتماعية الخاصة بك ، من المحتمل أن يكون لديك "عقدة" (أي دائرة في الخريطة) تمثل حساب Twitter أو صفحة Facebook. ومع ذلك ، فإن العلاقات بين هذه العقد (مثل علاقات الإعجاب، والمتابعة ، إعادة التغريد ، وما إلى ذلك) هي ما يمنح شبكة العقد ذات الصلة هيكلها. يمكنك التفكير في هذه العلاقة على أنها تحديد الهيكل العظمي للخريطة - من هذا الهيكل العظمي ، نستخلص العلاقات ذات الصلة بالعقد في الخريطة. في نظرية الرسم البياني - مجال علوم الكمبيوتر الذي يتعامل مع الشبكات ورسم الخرائط - يشار إلى هذه الاتصالات أيضًا بالحواف.

8 سبيل المثال ، قد يكون لنموذج الرسم البياني العدوى المعدية الفيروسية التي تمت مناقشتها أعلاه اتجاهًا مرتبطًا بانتشارها، وستنتقل الحافة من العارض إلى المصاب في هذه الحالة ، وسيتم تصويرها بسهم.

9 في مثال الرسم البياني للمطار أعلاه ، يمكن أن تكون قيم الحواف هي المسافة بين المطارات. قيمة حافة أخرى محتملة لهذا الرسم البياني هي الوقت الذي يستغرقه الطيران من مطار إلى آخر.

10 للحصول على أمثلة لشبكات المتابعين ، راجع شبكة الحسابات التي تروج للرسائل المضادة للقاحات في Wired هنا.

3. **تشذيب البيانات -** إذا قمت بجمع كمية كبيرة من البيانات ، فسينتهي بك الأمر دائمًا ببيانات أكثر مما يمكنك تعيينه. تميل أدوات Affinio و NodeXL و Gephi وأدوات إنشاء الخرائط الأخرى إلى العمل بشكل أفضل في نطاق عدة آلاف من العقد. عادةً ما يكون رسم الخرائط على نطاق أعلى مكلفًا جدًا من الناحية الحسابية بحيث لا يكون مفيدًا - ولهذا السبب ، من الضروري معرفة الشبكة الأكثر صلة بأسئلتك. غالبًا ما تسمى عملية قطع البيانات الدخيلة بالتقليم في علوم الكمبيوتر - قد يشار إليها أيضا باسم تقليل الشبكة أو تقليل الأبعاد. تقوم بعض الأدوات (مثل Graphika) بالتقليم نيابةً عنك ، ولكن يمكنك أيضًا اتخاذ قرارات بسيطة بنفسك باستخدام أدوات مثل Gephi. تتضمن بعض الأمثلة فقط عُقد الخرائط التي استخدمت علامة التصنيف ذات الصلة بالانتخابات أكثر من مرة ، أو تتضمن فقط العقد التي لها اتصال بخمس عُقد أو أكثر في الشبكة.[11]

4. **إنشاء الخرائط -** بعد تحديد العلاقة بالخريطة (الحواف) والعقد التي ستكون في شبكتك من خلال تقليل الشبكة ، فأنت على استعداد لإنشاء خريطتك. بشكل عام في هذه المرحلة ، سيكون عليك وضع بياناتك ذات الصلة في تنسيق يمكن قراءته للأداة التي تستخدمها (مثل ملف CSV أو ملف graphml. أو ملف .gexf لـ Gephi) ، وقراءتها في البرمجيات. بعد ذلك ، يتم تنفيذ معظم العمل الشاق من أجلك. هناك الكثير من البرامج التعليمية المجانية والمفيدة على YouTube لإنشاء خرائط الشبكة باستخدام Gephi[12] وأدوات أخرى مفتوحة المصدر.

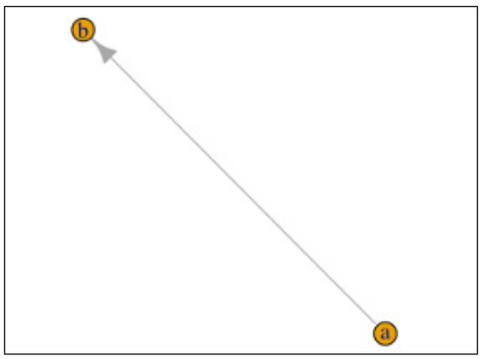
5. **تحليل الخرائط -** بمجرد إنشاء خريطتك ، يمكنك تخصيص الصور والانتقال إلى التحليل. عادةً ما تكون مقاييس المركزية أساسية لفهم التأثير في الشبكة. تحليل الشبكات الاجتماعية (Borgatti و Everett و Johnson 2019) لديه فصل كامل مخصص لأنواع مختلفة من المركزية التي تستحق النظر.[13]

أمثلة أنواع الشبكات

كما ذكر أعلاه ، هناك عدة أنواع من الشبكات التي يمكن إنشاؤها من مجموعة معينة من البيانات. يكمن العامل المحدد الرئيسي لنوع الشبكة التي تقوم بتعيينها في العلاقة التي اخترت تعيينها. على Twitter ، يُستخدم نوعان شائعان من الشبكات الموجهة بشكل متكرر لتحليل البيانات: شبكات المتابعة وشبكات إعادة التغريد. يمكن أن يكون كلا النوعين من الشبكات مفيدًا لتحليل البيانات والتأثير.

شبكات المتابعة:

في هذه الشبكة ، تمثل العقد حسابات Twitter وتمثل الاتصالات بينها العلاقات التالية. عادةً ما تكون هذه الاتصالات اتجاهية (تنتقل من عقدة إلى أخرى في اتجاه معين). يمكنك التفكير في هذه كخطوط تعني "يتبع". على سبيل المثال ، يوضح الرسم البياني أدناه عقدتين ، A و B ، ويوضح أن "A يتبع B".



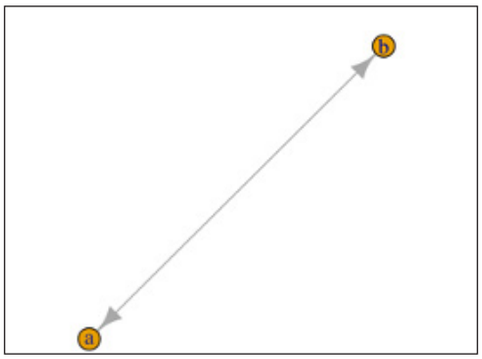
تصوير لعلاقة تتبع أحادية الاتجاه - في هذا الرسم البياني ،a يتبع b . وقد تم تصويره بخط موجه واحد (أو سهم) من a إلى b.

11 يُعرف هذا النوع من التقليم باسم تقليل النواة.

12 تقدم Gephi أيضًا برنامجًا تعليميًا مجانيًا بتنسيق PDF ومواد تعليمية أخرى على موقعها الرسمي - gephi.org.

13 الفصل 10 ، المركزية. هذا الكتاب هو نص رائع لتعلم أساسيات الشبكات وأساليب تحليل الشبكات. كما هو الحال دائمًا ، هناك أيضًا جميع أنواع البرامج التعليمية المجانية والمفتوحة المصدر الرائعة عبر الإنترنت.

لاحظ أنه في هذا الرسم البياني ، هناك فقط علاقة متابعة في اتجاه واحد - أي أن A يتبع B ، ولكن B لا يتبع A. إذا كانت هناك علاقة متابعة متبادلة ، حيث كان كل من A و B يتبعان بعضهما البعض ، سيظهر تصوير الشبكة الأسهم في كلا الاتجاهين ، كما في الشكل أدناه.



تصوير علاقة متبادلة متبادلة: A يتبع B ، B يتبع A.

في الشبكات التي يتم إنشاؤها على بيانات Twitter حول الانتخابات ، من المحتمل أن يكون لديك أكثر من عقدين في شبكتك ذات الصلة. تعمل الأمثلة أعلاه على إعطاء فكرة عن اللبنات الأساسية التي تم إنشاء شبكة معقدة منها.

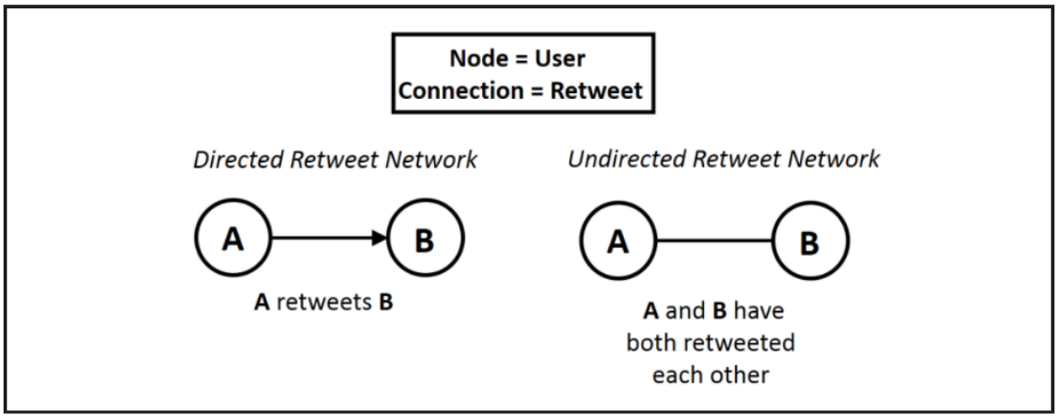
مميزات وعيوب شبكات المتابعة

العلاقات المتابعة على Twitter على وجه الخصوص هي علاقات طويلة الأمد ودائمة إلى حد ما - لا يلغي المستخدمون في الغالب متابعاتهم المستخدمين الآخرين. لهذا السبب ، تمثل الشبكات التالية رؤية طويلة المدى لديناميكيات الشبكة وتدفق المعلومات من شبكات إعادة التغريد. من ناحية أخرى ، يميل المستخدمون إلى أن يكون لديهم أكثر من اهتمام واحد على Twitter - وهذا يفتح احتمال أن بعض المستخدمين في شبكة المتابعين الخاصة بك قد لا تكون ذات صلة بالمحتوى الذي تهتم به كما تريد. غالبًا ما يكون للشبكات المؤتمتة مزيد من التركيز الموحد ، على سبيل المثال لدعم حزب معين أو مرشح أو مشكلة ، ولكن لا تزال غالبًا تختلف في المحتوى. إنه بالتأكيد جانب واحد يجب مراعاته عند تحليل حسابات التشغيل الآلي أو غيرها من النشاطات المنسقة.

قد يكون من الممكن تمامًا ، على سبيل المثال ، في خريطة افتراضية للطيف السياسي الأمريكي ، أن يكون هناك جزء من الشبكة التي تغرد في المقام الأول حول ثقافة البوب والموسيقى ، ولكنها كانت مرتبطة بشكل كبير بشبكة من الحسابات السياسية في المقام الأول. هذه المزايا والعيوب مفيدة لتضعها في الاعتبار عند تحديد ما إذا كانت شبكة المتابعين مهمة أم لا.

شبكات إعادة التغريد

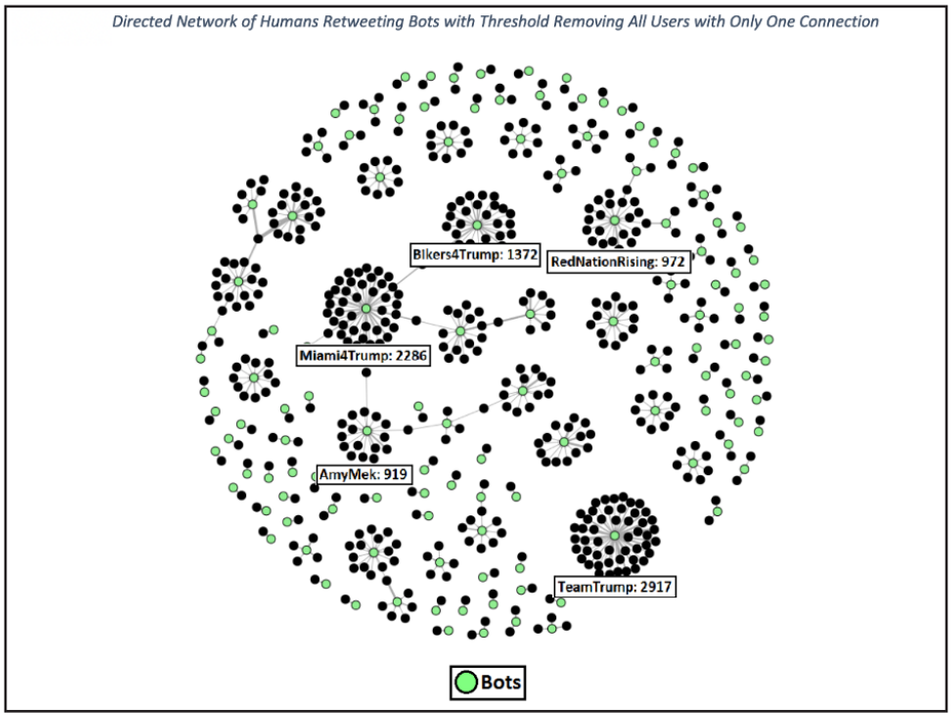
تعد شبكات إعادة التغريد نوعًا آخر من العلاقات على Twitter - تمثل الروابط بين العقد في هذه الشبكات الشخص الذي أعاد التغريد من ضمن البيانات التي تم جمعها. في هذا الصدد ، فإن هذه الخرائط موجهة أكثر نحو المحتوى من شبكات المتابعين. يظهر أدناه تصوير للحواف ضمن هذا النوع من الرسوم البيانية.



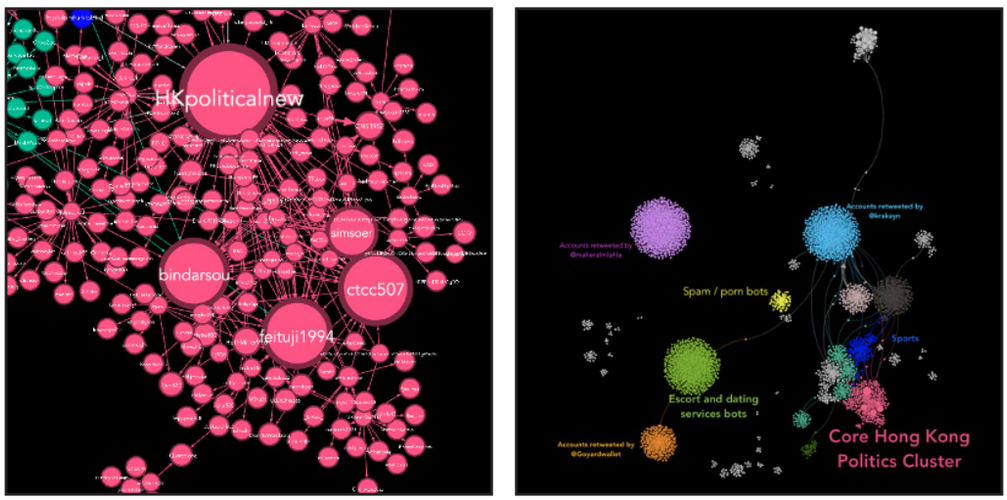
رسم تخطيطي يصور العقد والحواف داخل شبكة إعادة التغريد. (المصدر: Samuel Woolley and Douglas Guilbeault (2017). الدعاية الحسابية في الولايات المتحدة: إجماع التصنيع عبر الإنترنت. متوفر هنا.)

مميزات وعيوب شبكات إعادة التغريد

تعتبر هذه الشبكات سريعة الزوال مقارنة شبكات المتابعين ، حيث تمثل هذه الشبكات إعادة تغريد لقطة في في إطار زمني معين. في هذا الصدد ، فهي تصوير دقيق لديناميكيات التأثير في إطار زمني قصير - مثل حملة الهاشتاج المخصصة أو الأيام التي تسبق الانتخابات.



شبكة إعادة تغريد تم إنشاؤها من بيانات Twitter التي تم جمعها حول الانتخابات الرئاسية الأمريكية لعام 2016. في هذه الشبكة ، تكون العقد التي تمثل حسابات الروبوتات خضراء ، والحسابات البشرية سوداء. تصور شبكة إعادة التغريد هذه إعادة التغريد بين المستخدمين وتبين أن البشر أعادوا بشكل كبير إعادة تغريد محتوى الروبوت في الانتخابات الرئاسية لعام 2016. (المصدر: Samuel Woolley and Douglas Guilbeault (2017). الدعاية الحسابية في الولايات المتحدة: إجماع التصنيع عبر الإنترنت. متوفر هنا.)

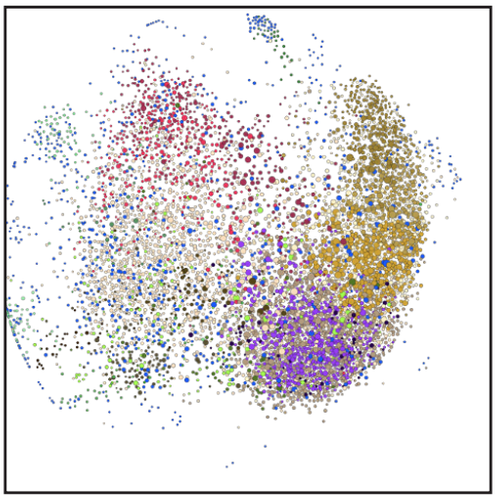


صور شبكة إعادة التغريد لحسابات الحكومة الصينية تنشر التضليل لتشويه سمعة الاحتجاجات المؤيدة للديمقراطية في هونغ كونغ من يونيو إلى أغسطس 2019. تم تجميع شبكة إعادة التغريد هذه من أرشيف عمليات المعلومات على Twitter[14]. تمثل هذه الصور شبكة إعادة التغريد بأكملها (يسار) وعرضًا مكبّرًا للمجموعة السياسية الرئيسية التي تستهدف هونغ كونغ (يمين). (المصدر: مختبر الذكاء الرقمي في تقرير معهد المستقبل ، متوفر هنا.)

14 على وجه التحديد من إصدار أغسطس 2019 للتغريدات المنسوبة للحكومة الصينية.

شبكات التنويه

تعد خرائط الإشارات نوعًا شائعًا آخر من الخرائط التي من المحتمل أن تواجهها. الفكرة هي نفسها تقريبًا مثل خرائط إعادة التغريد - تمثل الحواف "الإشارات" - الحالات التي يذكر فيها مستخدم آخر. باستخدام بيانات Gab التي تم تجميعها بواسطة Pushshift.io ، أنشأت Graphika خريطة شبكة إشارات لتحليل المحادثات والمجتمعات على المنصة الاجتماعية Gab - وهي شبكة اجتماعية تشبه إلى حد كبير موقع Twitter في استخدامه والبناء.



خريطة الرسم البياني لمستخدمي Gab. تم جمع هذه الخريطة من مجموعة بيانات Gabshift.io المتاحة للجمهور ، للفترة من أغسطس 2016 إلى أواخر أكتوبر 2018.

شبكات المتابعة ، وإعادة التغريد ، والإشارة ما هي إلا مجموعة من الخيارات المتاحة لك لإنشاء شبكات من بيانات Twitter - هناك بالتأكيد إمكانيات أخرى. على منصات أخرى ، مثل Facebook أو Gab ، قد تكون المجالات الأخرى ممكنة. يمكن للمرء أن يتخيل خريطة شبكة من الصفحات العامة تجمعها الاعجابات المتبادلة، على سبيل المثال.

قيود جمع البيانات وتخطيط الشبكة

كما هو الحال مع أي أداة ، فإن تعيين الشبكة له قيود معينة ، وكذلك عملية جمع البيانات التي تسبقها. من أهم القيود التي يجب وضعها في الاعتبار النطاق والوقت. كما هو مذكور في قسم جمع البيانات ، فإن جميع الأدوات المتاحة لإنشاء خرائط الشبكة لديها حاليًا قيود على المقياس ، سواء كانت مجانية أو مدفوعة. هذا لأنه مع نمو عدد العقد في الشبكة ، ينمو عدد الاتصالات المحتملة بشكل كبير[15]. هذه الكمية من البيانات تفرض ضرائب ، حتى بالنسبة لأجهزة الكمبيوتر ، وبالتالي من غير المحتمل أن تجد أداة يمكنها إنشاء خريطة أكبر من حوالي 15000 عقدة[16]. سيكون من الخطأ رؤية هذا القيد على وجه الحصر على أنه مشكلة ، ومع ذلك ، فإن جزءًا مهمًا من علم البيانات وتحليلها ينطوي على اختيار أجزاء مجموعة البيانات الكبيرة التي من المرجح أن تقدم رؤى مفيدة. بعبارة أخرى ، يتضمن تحليل البيانات بشكل صحيح اختيار أجزاء البيانات الأكثر جدارة بالتحقيق - وبهذه الطريقة ، يعتبر علم البيانات فنًا بقدر ما هو علم. هذه المهمة ، التي تعد جزءًا مركزيًا من فحص البيانات وتحليلها ليست بالأمر السهل, فيصعب الوصول إلى بيانات أكثر.

القيد الثاني الذي من المحتمل أن تواجهه مع جمع البيانات ورسم خرائط الشبكة هو الإطار الزمني ، خاصة عند جمع البيانات التاريخية. تسمح واجهة برمجة تطبيقات بحث Twitter القياسية بجمع البيانات التاريخية فقط بعد أسبوع من وقت الاستعلام. يجب أن تأتي البيانات الأقدم من 9-7 أيام من مصدر آخر - مثل أداة (مدفوعة) تتيح جمعًا تاريخيًا بعيد المدى ، أو شراء بيانات من مزود بيانات ، مثل GNIP. في حين أن تحليل البيانات التاريخية يمكن أن يكون مفيدًا في حالات معينة ، فإن شراء البيانات التاريخية يمكن أن يصبح مكلفًا بسرعة. لهذا السبب ، **فإن أفضل حل هو دائمًا جمع جميع البيانات ذات الصلة فور حدوثها.** إذا كنت أنت أو أحد أعضاء فريقك مهتمًا ببث أو جمع البيانات حول الانتخابات ، فمن الأفضل دائمًا البدء في الجمع بمجرد أن يكون لديك فكرة واضحة عما ستهتم به.

15 بالنسبة للمهتمين بالرياضيات ، فإن شبكة العقد تحتوي على n * (n-1) / 2 اتصالات محتملة.

16 برنامج رسم خرائط Graphika قادر حاليًا على رسم الخرائط على أوسع نطاق ، مما يتيح رؤية ما يصل إلى 5.5 مليون عقدة

الشبكات المغلقة والمشفرة

يتمثل التحدي الذي يواجه جهود مراقبة وسائل الإعلام في السنوات الأخيرة في الشعبية والتطبيق الواسع النطاق لتطبيقات الرسائل المشفرة مثل WhatsApp و Telegram و Signal. بشكل متزايد ، تحدث وسائل الإعلام التي يعتمد عليها المواطنون لمواكبة التطورات السياسية خلال الانتخابات وغيرها من الأحداث السياسية الرئيسية على هذه المنصات. شهدت دول مثل البرازيل والهند والمكسيك زيادة ملحوظة في الرسائل السياسية على WhatsApp ، على سبيل المثال. في حين أن استخدام شبكات الرسائل المشفرة يعد بلا شك مفيدًا للخصوصية والأمن والحقوق الرقمية للمواطنين ، فإنه يفرض تحديات جديدة لفهم نشر المعلومات السياسية عبر الإنترنت.

في الوقت الحاضر ، تعتمد أفضل الطرق لجهود مراقبة وسائل الإعلام للشبكات المغلقة على التحقق اليدوي للحقائق - فرق الخبراء التي تراقب قنوات WhatsApp ذات الصلة بشكل فردي أو كمجموعة ، تحقق من القصص وتوزع نتائج هذه الجهود علنًا.كانت مثل هذه الجهود ناجحة في العديد من الحالات مثل كاشف La Silla Vacía WhatsApp[17] في كولومبيا ، وكذلك العمل من قبل Verificado[18] في المكسيك ، ومركز الديمقراطية والتنمية في نيجيريا[19].

جهد آخر جدير بالملاحظة هو بوت كوفاكتس في تايوان. ابتكر جونسون لياغ وفريق من المطورين العاملين مع حركة gOv طريقة للجمع بين التحقق اليدوي من الحقائق والتوزيع الآلي لمساعدة المواطنين التايوانيين على التحقق مما إذا كانت القصة كاذبة أم لا. يمكن للمستخدمين إضافة روبوت Cofacts[20] على LINE ، وهو تطبيق مراسلة مشفر شائع يستخدم في تايوان واليابان. إذا شاهد المستخدم قصة مشبوهة ، فيمكنه لصق الرابط في محادثة إلى روبوت Cofacts. إذا لم تتم مشاهدة القصة من قبل ، يقوم فريق من البشر بالتحقق من القصة وتحميل رسالة تلخص صحة القصة في قاعدة بيانات مركزية. يقوم برنامج التتبّع بعد ذلك بإعادة توجيه هذه الرسالة إلى المستخدم الأصلي وأي مستخدمين آخرين لديهم فضول لمعرفة ما إذا كانت القصة صحيحة أم لا. يتيح Cofacts الوصول العام إلى البيانات المجهولة التي جمعتها على Github ، كما يسمح للمستخدمين بالبحث في قاعدة البيانات هذه من خلال موقع ويب عام. أصدرت Meedan ، وهي شركة تدعم التحقق من الحقائق وغيرها من الأبحاث عبر الإنترنت ، مجموعة مماثلة من الأدوات على Check تساعد مختلف المستخدمين على أتمتة وإدارة سير العمل بشكل جماعي لعملية التحقق من الحقائق على WhatsApp والمنصات الأخرى.

توجد بعض الأدوات ، مثل - النسخ الاحتياطي للمحادثات في تطبيق WhatsApp- Backup WhatsApp Chats ، والتي تسمح للمستخدمين بتصدير محادثات WhatsApp إلى ملفات CSV ، ولكن لا يزال يتعين على المستخدم أن يكون مضموما في هذه الدردشات . Telegram ، تطبيق مراسلة مشفر ، لديه واجهة برمجة تطبيقات تسمح للمستخدمين بالوصول إلى القنوات العامة برمجيًا. هذا يسمح بقدر من مراقبة وسائل الإعلام العامة ، على الرغم من أن البيانات ليست غنية مثل تويتر.

أدوات وحزم للتصور الشبكي المفيد

يسرد الجدول أدناه العديد من الأدوات الشائعة وحزم الكود التي يتم استخدامها لتصوروتحليل الشبكة.

أدوات التحليل والتصور الشبكي			
أدوات مفتوحة المصدر / مجانية	أدوات مدفوعة	الحزم شائعة الاستخدام (Python)	الحزم شائعة الاستخدام (R)
Gephi NodeXL	Graphika Affinio	networkx matplotlib igraph	igraph plotrix

17 https://www.niemanlab.org/2017/03/to-slow-the-spread-of-false-stories-on-whatsapp-this-colombian-news-site-is-enlisting-its-own-readers/

18 https://www.niemanlab.org/2018/06/whatsapp-is-a-black-box-for-fake-news-verificado-2018-is-making-real-progress-fixing-that/

19 https://www.cddwestafrica.org/whatsapp-nigeria-2019-press-release/

20 الاسم الصيني لهذا الروبوت (الصواب والخطأ) 真的假的

تحديد المؤثرين والمجموعات والحسابات

الهدف من جمع البيانات وتحليلها هو فهم كيفية توزيع المعلومات داخل الشبكة. ما هي القصص الأكثر شعبية؟ من هم المستخدمون الأكثر تأثيرًا؟ ما هي المجالات الإخبارية التي يتم الاستشهاد بها كثيرًا في المحادثة؟ باستخدام مجموعة بيانات قوية والأدوات المناسبة ، يمكنك البدء في الإجابة عن هذه الأسئلة بخصوصية واستنباط ديناميكيات نشر المعلومات في المساحة عبر الإنترنت التي تراقبها.

هناك طريقتان لتصور التأثير على وسائل التواصل الاجتماعي: يمكننا تسمية هذه الأساليب القائمة على المحتوى والفاعلية لتحديد التأثير. سنستكشف كلاهما أدناه.

طرق قائمة على المحتوى لتحديد التأثير

تركز هذهالأساليبعلى التغريدات أو الوسمات للكلمات الرئيسية أو مواقع الويب في شكل عناوين URL أو النطاقات[21]. في بعض السيناريوهات ، سيتم معرفة الممثلين الذين يستحقون المشاهدة - مثل المنافذ الاعلامية أو السياسيين الذين ينشرون المعلومات المضللة بشكل متكرر ، على سبيل المثال. من ناحية أخرى ، من الشائع جدًا عدم معرفة مصادر التضليل أو الكلام الذي يحض على الكراهية أو الأشكال الأخرى من المحتوى الذي تبحث عنه.

غالبًا ما يكون مفيدًا للغاية عند تحليل المحادثات عبر الإنترنت حول الانتخابات أو الخطاب السياسي الآخر لفهم المحتوى الذي يحصل على أكبر قدر من التفاعل - خاصة عندما لا يكون هناك ممثل معين تتطلع إلى تحليله والذي حددته في البيانات. يمكن أن يكون النظر إلى عناوين URL أو التغريدات التي تحصل على أكبر قدر من التفاعل في البيانات نطة بداية جيدة لهذا السيناريو.

التغريدات

عند فحص التغريدات ، يمكنك مقاربة تأثيرها من خلال النظر في عدد إعادة التغريد و / أو الإعجابات التي حصلت عليها[22]. سواء كنت تستخدم أداة خارجية أو تسحب البيانات من Twitter API ، يجب أن يكون لديك وصول سهل إلى هذه البيانات في جميع الأوقات. بعد تحديد التغريدات الأكثر تأثيرًا ، يمكنك استخدام هذه التغريدة كنقطة انطلاق لمزيد من التحقيق. تتضمن بعض الأسئلة التي تستحق التحقيق ما يلي:

- من صاحب التغريدة في الأصل؟ من قام بإعادة تغريد المنشور؟ كيف تبدو متابعات هؤلاء المستخدمين؟ إذا كانت كبيرة ، فقد يكون من المفيد إجراء تحليل شبكة لها.
- ما هي علامات الوسم المستخدمة في التغريدة؟ إذا كان أي منهم مميزًا أو مدفوعًا فقط من قبل مجموعة صغيرة من المستخدمين ، فهل هناك شيء مشترك بين هؤلاء المستخدمين؟
- ما عناوين URL الموجودة في المنشور؟ إذا كان عنوان URL مشبوهًا (تم إنشاؤه مؤخرًا أو يشجع على التضليل) ، فقد يستحق الأمر مزيدًا من التحقيق الإضافي - التحقق من سجلات تسجيل النطاق من خلال بحث Whois[23] ، أو البحث في Twitter عن إشارات أخرى مثيرة للاهتمام لعنوان URL ، على سبيل المثال. يمكنك أيضًا استخدام امتداد متصفح CrowdTangle لمعرفة ما إذا كان عنوان URL هذا يكتسب تفاعل على Facebook أو Instagram أو Reddit أو في أي مكان آخر على Twitter.

تنطبق هذه الاستراتيجيات والمبادئ نفسها أيضًا على Facebook و Twitter و Gab وغيرها من منصات وسائل التواصل الاجتماعي - يعد تحليل التفاعلات مع المنشور طريقة موثوقة لقياس تأثير الرسالة في بيئة معينة.

علامات الوسم / الكلمات الرئيسية

علامات الوس هي بطبيعة الحال واحدة من الكيانات الرئيسية ذات الأهمية عند فحص Twitter والبيانات الاجتماعية الأخرى. إن استخدام علامة الوسم لتسليط الضوء على موضوع ما يتم نشره على تويتر هو نعمة للباحثين - فهو يمكّننا من جمع بيانات المحادثة ذات الصلة بموضوع محل اهتمام بسهولة كبيرة. فبمجرد تحديد علامة الوسم أو مجموعة منها ذات الأهمية ، تفتح المجال لتحقيق أعمق على سبيل الذكر لا الحصر: تحليل علامات الوسم الأكثر شيوعًا ، أو تقسيم اقتباسات علامة الوسم حسب الوقت ، أو تحليل عناوين URL المتزامنة.

21 تشير عناوين URL إلى ارتباط كامل ينقلك إلى قصة أو موقع معين ، بينما تشير المجالات إلى الموقع الذي يستضيف هذا المحتوى - وهذا يتوافق مع النص الذي يسبق نطاق المستوى الأعلى (يُختصر باسم TLD - مثل .gov ، com ، org ، إلخ.) و TLD نفسها. على سبيل المثال ، تتم استضافة عناوين URL الثلاثة example-news-site.com/story1 و example-news-site.com/story2 و example-news-site.com/story3 على نفس النطاق - example-news-site.com.

22 ملاحظة مثيرة للاهتمام تذكرها للتغريدات هي أن التغريدات تحصل على عدد أكبر من الإعجابات من إعادة التغريد. هذا مشابه لحقيقة أن معظم مشاركات Facebook لديها إعجابات أكثر من المشاركات. إذا كانت هذه النسب غير طبيعية ، يمكن أن تكون مؤشرًا على نشاط غير حقيقي ، على الرغم من عدم وجود ضمان.

23 هناك العديد من قواعد بيانات Whois للتحقق من تفاصيل التسجيل ، واحدة موثوقة تحتفظ بها شركة الإنترنت للأسماء والأرقام المخصصة (ICANN) https://lookup.icann.org/

عناوين URL / النطاقات

تعد عناوين URL والنطاقات التي تحدث داخل التغريدات ، والتي غالبًا ما ترتبط بمصادر الأخبار ، مصدرًا مفيدًا للغاية لتحديد المحتوى والمنشورات والسرد الذي يحصل على أكبر قدر من التفاعل داخل مجتمع معين عبر الإنترنت.

تتمثل الخطوة الأولى الشائعة لتحليل عنوان URL / النطاق في استخراج عناوين URL الفريدة من مجموعة بيانات وحساب عدد المرات التي يتم فيها الاستشهاد بها في المجموعة. هذه التقنية بسيطة وقوية ، غير أن من السهل أن تخطئ بشكل منطقي في أكثر النقاط دقة. للتأكد من أن التحليل الخاص بك يقدم أفضل الأفكار وأكثرها دقة ، يجدر الانتباه إلى عدد قليل من المشكلات أدناه.

- **حل عناوين URL المختصرة -** غالبًا ما يتم اختصار عناوين URL في التغريدات لتلائم حدود الأحرف - استخدام أداة حل عناوين URL يمنحك عنوان URL الكامل الذي يشير إليه عنوان URL المختصر. توفر بعض الأدوات ، مثل URLex.org ، واجهة برمجة تطبيقات للحل المجمّع والآلي لعناوين URL. هذه هي أفضل طريقة للتأكد من عدم فقد البيانات في عناوين URL المختصرة.

- **توحيد عناوين URL -** يمكن الاستشهاد بنفس النطاق باستخدام سلاسل نصية مختلفة[24]. على سبيل المثال ، قد تحدث روابط إلى New York Times في نص مثل www.nytimes.com و nytimes.com و NYTimes.com و http://nytimes.com و nyti.ms و https://nytimes.com و http: //www.nytimes.com أو https://www.nytimes.com أو m.nytimes.com - هذه ليست سوى عدد قليل من الاحتمالات. من أجل ضمان عدم الاستهانة بتأثير عنوان URL أو النطاق من خلال حساب السلاسل المختلفة ، ستحتاج إلى التأكد من توحيد التنسيق قبل العد. لن يكون من الممكن دائمًا التحكم في جميع الطرق المتنوعة التي يمكن أن تحدث بها عناوين URL المرتبطة بالمحتوى نفسه ، ولكن القليل من التفكير في التوحيد القياسي في وقت مبكر يمكن أن يقطع شوطًا طويلاً نحو تعزيز تحليلاتك. الأشياء التي يجب مراعاتها عند التوحيد هي (1) الحساسية لحالة الأحرف[25] ، (2) البادئات (ww2 ، www. ، https: // ، http: // ، إلخ.) ، و (3) النطاقات الفرعية ، من بين عوامل أخرى.

بعد تحليل عناوين URL الأكثر شيوعًا في مجموعة ، لديك العديد من الخيارات لمزيد من التحليل. إذا كان النطاق مصدرًا جديدًا للأخبار نسبيًا على المشهد ، يمكنك استخدام تقنيات الذكاء المفتوح المصدر (OSINT) مثل التحقق من معلومات تسجيل النطاق للبدء في اكتساب نظرة ثاقبة حول اتصالات النطاق. OSINT هي طريقة مفيدة لجمع المزيد من المعلومات حول الحسابات أو مواقع الويب ذات الأهمية. إنه حقل متغير باستمرار ، حيث تظهر الأدوات والتقنيات وتختفي كل يوم. تعد Intel Techniques من أفضل المصادر لتعلم مهارات جديدة ، ولكن هناك العديد من المصادر الأخرى. قامت Bellingcat ، وهي مجموعة من الباحثين التي غالبًا ما توثق عمليات التأثير الروسية ، بعمل فني باستخدام OSINT لتقديم قصص صحفية رائدة ، وتحتفظ بمستند[26] عام من Google مع قائمة شاملة بأدوات التحقيق وأدوات OSINT. استخدم لورنس ألكسندر ، عالم البيانات المقيم في المملكة المتحدة ، ببراعة رموز Google Analytics لرسم شبكة من مواقع الويب المؤيدة للكرملين في عام 2015.

بعد تلخيص المواقع والمقالات الأكثر تأثيرًا في مجموعة بيانات معينة ، يمكنك أيضًا إجراء بعض تحليل المحتوى لتحليل السرد الذي يحدث في هذه المقالات. تقنيات NLP (معالجة اللغات الطبيعية) مثل استخدام ترددات n-gram لتحليل العبارات الأكثر شيوعًا في هذه المقالات ، tf-idf لمقارنة الموضوعات النسبية لمقالات مختلفة ،أو استخدام طرق تحليل المحتوى النوعي يمكن أن تؤدي جميعها إلى نتائج إخبارية بمجرد تضييق نطاق بعض المقالات وعناوين URL المؤثرة.

24 غالبًا ما يشار إلى البيانات النصية في سياق حسابي باسم "سلاسل". هذا هو اختصار لسلسلة من الشخصيات ، وهذه هي الطريقة التي سيشير بها علماء الكمبيوتر غالبًا إلى نص التغريد / النشر في مجموعات بيانات وسائل التواصل الاجتماعي.

25 تشير حساسية حالة الأحرف إلى ما إذا كانت البيانات النصية صغيرة أو كبيرة. تتم معالجة هذه المشكلات بسهولة أكبر من خلال تقليل عدد عناوين URL الموجودة في المجموعة قبل حساب عدد مرات تكرار كل عنوان URL. من المهم ملاحظة أنه على الرغم من أن معظم عناوين URL الكاملة ليست حساسة لحالة الأحرف ، إلا أن العديد من عناوين URL المختصرة تكون. لذلك يوصى بأن يعالج الباحثون أولاً دقة عنوان URL قبل الانتقال إلى عناوين URL ذات الأحرف الصغيرة / المعيارية.

26 https://docs.google.com/document/u/1/d/1BfLPJpRtyq4RFtHJoNpvWQjmGnyVkfE2HYoICKOGguA/edit

طرق قائمة على الشبكة

هناك طريقة أخرى لتحليل التأثير في مجموعة البيانات وهي اتباع نهج قائم على الشبكة. في هذا النهج ، تستخدم بياناتك لبناء شبكة ذات صلة كما نوقش في قسم تحليل البيانات والشبكة أعلاه - قد يكون هذا عبارة عن إشارات أو متابعين أو إعادة تغريد أو مقاييس أخرى تستند إلى الشبكة.

المجموعات / التكتلات

يعد اكتشاف المجتمعات أساس العمل الأكثر أهمية الذي يقارن المجتمعات (غالبًا ما يشار إليها بالمجموعات في نظام الحسابات القومية). في حين أن هناك نظريات منهجية معقدة تكمن وراء خوارزميات مختلفة للتجميع ، فإن معظم العمل الشاق يتم من أجلك في أي برنامج ستستخدمه. تقدم Gephi العديد من الخيارات للتخطيط والتجميع (يمكنك العثور على مزيد من التفاصيل في البرنامج التعليمي هنا). كما تقوم Graphika بهذا العمل تلقائيًا. يكفي أن نقول أنه ، بمجرد أن يحدد برنامج التصور والتحليل الشبكي عدد المجتمعات المختلفة في شبكتك ، يمكنك الانتقال إلى التحليل النوعي لتحديد ما هو مشترك بين أعضاء مجموعة معينة.

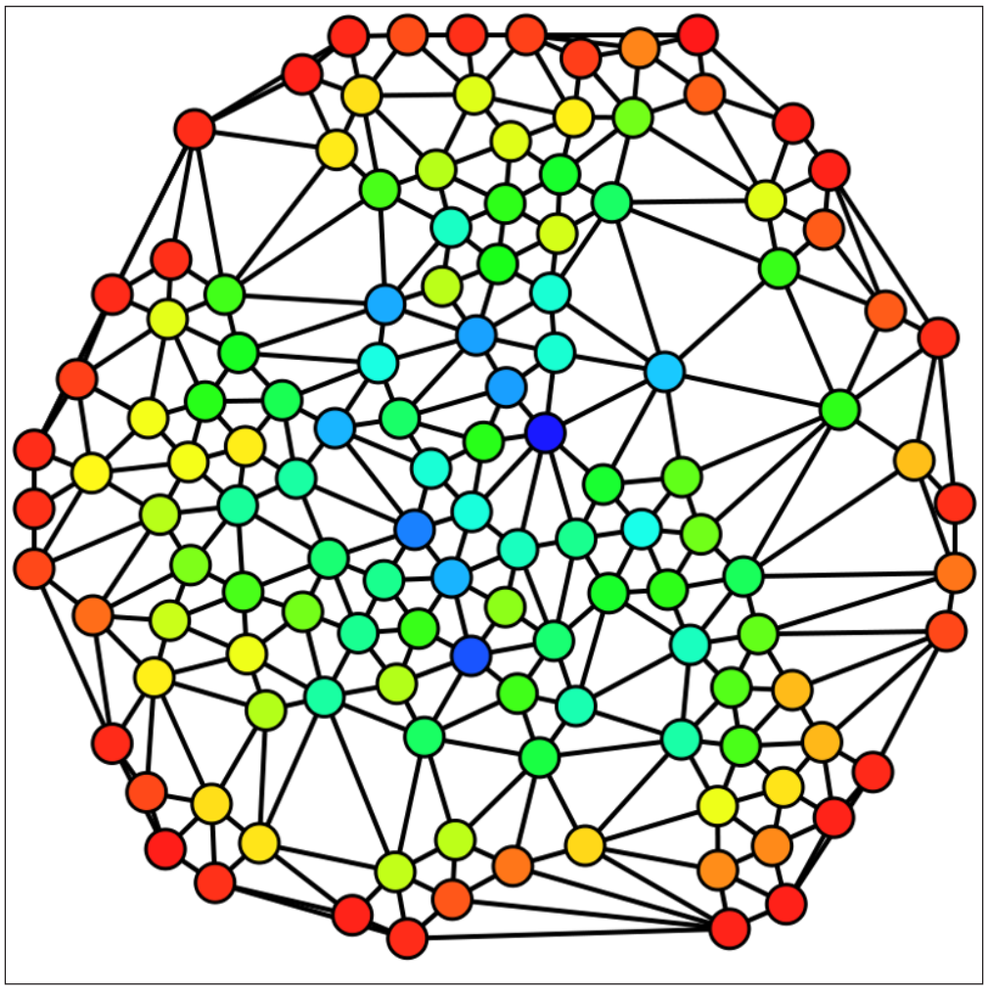
في كثير من الأحيان ، سيكون للمجتمعات ومجموعات الحسابات ميزات مشتركة يمكن تمييزها ، مثل الترويج لمصادر إخبارية مماثلة أو الانتماء إلى طرف مشابه. هنا يصبح فهم السياق السياسي العام مهمًا. إن الجمع بين التحليل الكمي لبيانات المجتمع وتحليل المحتوى النوعي هو أفضل طريقة لتحديد ما هو مشترك بين أعضاء مجتمع معين.

المركزية

في نظرية الشبكة ، على وجه التحديد كما ينطبق على نظام الحسابات القومية ، فإن المقياس الرئيسي لتحليل تأثير ممثل واحد أو مجموعة من الجهات الفاعلة هو المركزية. المركزية هي مقياس لكيفية اتصال العقدة المحددة أو مدى تأثيرها داخل الشبكة. هناك عدة طرق مختلفة لقياس المركزية.

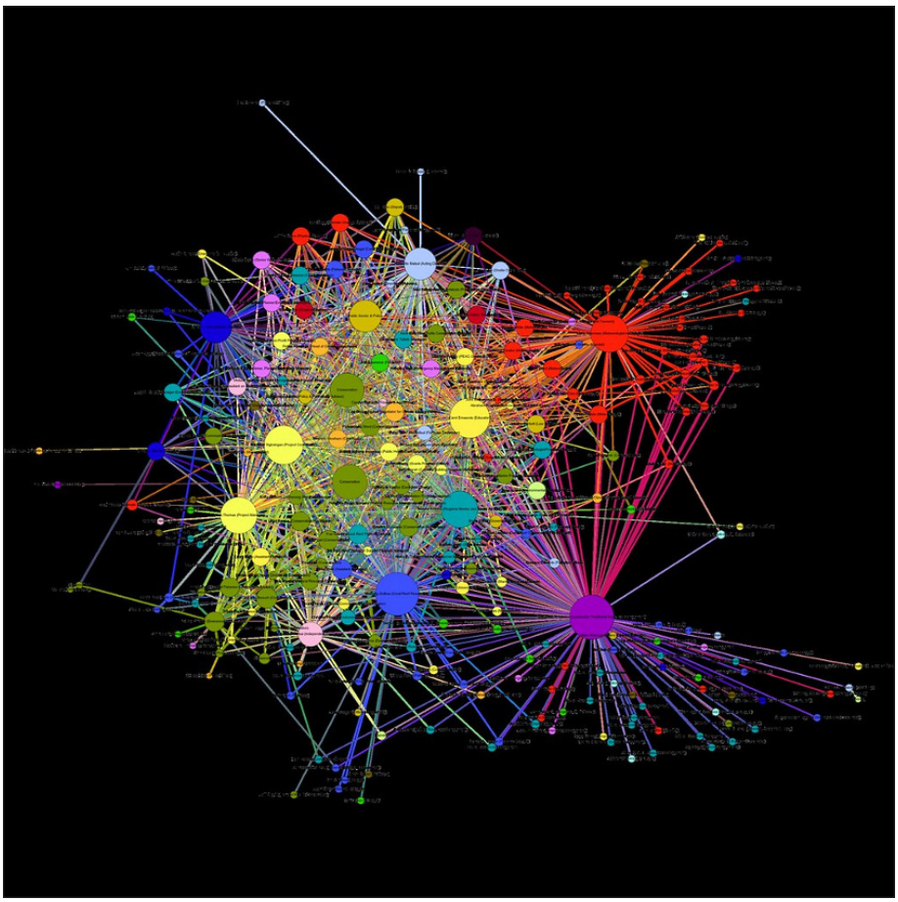
- **مركزية الدرجة -** في الرسم البياني غير الموجه ، تكون مركزية الدرجة لعقدة معينة هي عدد الاتصالات المباشرة التي لديها بالعقد الأخرى حولها. في الرسم البياني الموجه ، حيث يؤخذ اتجاه الحافة في الاعتبار ، هناك نوعان من مركزية الدرجة - المركزية في الدرجة ، أو عدد الاتصالات الواردة للعقدة - ومركزية خارج الدرجة ، أو عدد الاتصالات الصادرة للعقدة. يمكن القول أن المركزية في هذه السيناريوهات هي الأكثر صلة - حيث أن العقدة التي تحتوي على عدد كبير من الاتصالات الواردة لها إمكانات عالية التأثير. هذا مفهوم بديهي نسبيًا - يتمتع الحساب الذي يحتوي على 15000 متابع (بدرجة 15000) بقدرة أكبر على التأثير بشكل مباشر من حساب لديه 100 متابع. ستختلف أهمية الدرجات الداخلية / الخارجية اعتمادًا على الشبكة التي تفحصها.

- **مركزية التداخل -** يمكن اعتبار مركزية التباين بمثابة قياس قدرة العقدة على نشر رسالة إلى العقد الأخرى بسرعة. بعبارة أخرى ، تمنح المركزية بين النقاط درجات أعلى للعقد التي من المحتمل أن تكون نواقل فيروسية أو نشر سريع للرسالة. يتم قياس هذه القيمة من خلال حساب عدد المرات التي تجد فيها العقدة نفسها في جزء من أقصر مسار من عقدة إلى عقدة أخرى.



تدل الصبغة (من الأحمر = 0 إلى الأزرق = الحد الأقصى) على مركزية كل عقدة. (كلاوديو روتشيني) المشاع الإبداعي ب 2.5
https://en.wikipedia.org/wiki/Social_network_analysis#/media/File:Graph_betweenness.svg

- **مركزية المتجه الذاتي -** تعتبر مفيدة جدًا لفهم التأثير داخل شبكة رئيسية أو شبكة فرعية معينة. الفكرة الكامنة وراء مركزية المتجه الذاتي هي في الأساس "من يؤثر على المؤثرين؟". مع مركزية المتجه الذاتي ، يتعلق التأثير الأقل بعدد العقد المستهدفة والمرتبطة بالعقد الأخرى ، والتأثيرالأكثر يتعلق بعدد الروابط للعقدة الهدف وارتباطها بالعقد الأخرى المتصلة جيدًا. مكنك التفكير في هذا الإصدار من المركزية على أنه نوع من "التأثير المتدرج" - ما مدى احتمالية انتشار المعلومات المتدفقة من عقدة المصدر إلى العقد المؤثرة الأخرى؟ يستخدم محرك بحث Google شكلاً من هذا المفهوم حيث أنه يصنف الصفحات (أو العقد) وفقًا لعدد الصفحات الأخرى التي تشير إليها. يمكن أن تلعب العقد مركزية المتجه الذاتي دورًا قويًا بشكل خاص في نشر البضائع (مثل المال والمعلومات والجراثيم وما إلى ذلك) من خلال شبكة.



مشروع تحليل الشبكات الاجتماعية الإقليمية المتكاملة وتقييمات الشبكة الاجتماعية ، خريطة تتمحور حول بالاو ؛ جميع الأعضاء من بالاو أو متصلون بها. يتم تحديد حجم العقد وفقًا لمركزية المتجه الذاتي وملونة وفقًا للمهنة.
https://www.flickr.com/photos/pacificrisa/11344578486 المشاع الإبداعي بمعدل 2.5

- **الجسور -** على الرغم من أنها ليست شكلاً من أشكال المركزية ، تعد الجسور ظاهرة مفيدة في بنية الشبكة يجب وضعها في الاعتبار. تعمل العقدة أو مجموعة صغيرة من العقد كجسر عندما تقوم بتوصيل مجموعة كثيفة بمجموعة كثيفة أخرى. بدون الجسور ، لن يكون لمجموعات المصالح المتنوعة (مثل أعضاء حزبين سياسيين منفصلين ، أو هواة السياسة في بلدين مختلفين) أي وسيلة لنقل المعلومات أو التواصل بين بعضهم البعض. في هذا الصدد ، تعد الجسور أجزاء مهمة من المحتوى الفيروسي على الشبكات عبر الإنترنت.

- **الفروع الخارجية -** عبارة عن عقد في رسم بياني مع حافة واحدة فقط تربطها بشبكة. غالبًا ما تستحق هذه الحسابات الإزالة من الشبكة ، خاصة الشبكات الكبيرة ، حيث من غير المرجح أن يكون لها أي تأثير.

تتوفر العديد من البرامج التعليمية والموارد المفيدة عبر الإنترنت لقراءة أنواع مختلفة من المركزية - أحد الموارد المفيدة هو كتاب تحليل الشبكات الاجتماعية (Borgatti et al. , 2013) ، الذي يحدد أنواعًا مختلفة من المركزية وكيف يتم قياسها في الفصل 10.

يمكن للأدوات مفتوحة المصدر مثل Gephi القيام بالحسابات الصعبة لحساب أشكال مختلفة من المركزية بالنسبة لك في ثوانٍ[27]. إن اختيار الشكل المركزي للاستخدام هو في النهاية اختيار يقع على عاتقك أنت وفريقك. تعد استشارة عالم البيانات أو منظّر الشبكة الخيار الأمثل ، لا تحتاج إلى التأكيد كثيرًا على التفاصيل المنهجية هنا - تفترض جميع أشكال المركزية تقريبًا أن العقد الأكثر تأثيرًا هي ، بمعنى ما ، العقد ذات أكبر عدد من الروابط في شبكة رئيسية أو شبكة فرعية.

بمجرد تحديد شكل المركزية التي ستستخدمها ، يمكنك تخصيص مرئيات شبكتك لتعكس ذلك. يتيح لك Gephi استخدام حجم العقدة أو ظلال الألوان لإبراز المركزية ، على سبيل المثال.

27 أو دقائق ، حسب حجم الشبكة المعنية.

تحديد مصادر الأخبار

بالنسبة للانتخابات وتحليل المعلومات على وجه الخصوص ، فإن مسألة عدد مصادر الأخبار المذكورة في البيانات ذات أهمية مركزية. هذا مجال يكون فيه من المفيد بشكل خاص العمل مع الأفراد الذين لديهم فهم اجتماعي وسياسي شامل للبلد أو المنطقة المعنية. غالبًا ما تكون مصادر الأخبار التي يتم الاستشهاد ناتجة عن معرفة مسبقة لهؤلاء الأفراد. من المهم أن يكون لديك فهم قوي لبيئة وسائل الإعلام الأساسية في الدولة ، كما هو موضح أعلاه ، بما في ذلك حصة وتأثير سوق التلفزيون والإذاعة والصحف ووسائل الإعلام عبر الإنترنت.

هناك العديد من الطرق المفيدة لتحديد مصادر الأخبار الجديدة في منطقة أو مجموعة بيانات. أحد الخيارات هو إجراء عمليات بحث يدوية متكررة من Google للأخبار والسياسة في المنطقة - خاصةً بلغات متعددة إذا كانت المنطقة متعددة اللغات. من المرجح أن تتمتع مصادر الأخبار التي تتنافس من أجل التأثير بتحسين قوي لمحركات البحث (SEO) لإظهارها ضمن النتائج الأولى التي تحتفظ بها Google ، ويمكنها غالبًا عرض مصادر الأخبار التي تم إنشاؤها مؤخرًا والتي لا يعرفها المواطن العادي. هناك خيار آخر ، يوصى به بشدة ، وهو إجراء تحليل الاقتباس للنطاق وعنوان URL على مجموعة بيانات. يمكن أن يتخذ هذا بعض الأشكال - مراقبة CrowdTangle لعناوين URL الرائدة في مجموعة معينة من الصفحات السياسية ، أو استخراج عناوين URL / نطاقات فريدة من بيانات النشر[28] في مجموعة بيانات مجمعة مع خبير محلي والبحث عن النطاقات التي لم ترها من قبل. يحتوي Crowdtangle على مكوّن إضافي مجاني لمتصفح Chrome والمتصفحات الأخرى يسمح لك بمشاهدة المشاركات لمقالة فردية ، ولكن للوصول إلى لوحة المعلومات الأكثر تقدمًا والأدوات الأخرى ، يجب على المستخدمين الحصول على ترخيص مؤسسي ، غالبًا عبر Facebook ،المالك له الآن.

الفكرة التوجيهية هنا هي أنه غالبًا ما ينبثق التضليل من النطاقات التي تم إنشاؤها حديثًا والتي تظهر بسرعة حول القضايا السياسية والانتخابات ، وتختفي بنفس السرعة. كما الحال مع streetnews [.] واحد ، وهو النطاق الذي تم إنشاؤه والترويج له في Gab في الفترة السابقة خلال الانتخابات النصفية في أمريكا لعام 2018. تم استخدام هذا النطاق للترويج لمحتوى كراهية الإسلام والأخبار المثيرة قبل أن تختفي بسرعة. كان هذا هو الحال أيضًا في Endless Mayfly ، وهو تحليل لمعلومات إيرانية مضللة أجراها مختبر سيتيزن بجامعة تورونتو جنبًا إلى جنب مع خبراء الصناعة. في هذه الحالة ، صاغ فريق Citizen Lab عبارة "التضليل السريع الزوال" لوصف نطاقات التضليل التي تظهر بسرعة وتختفي التي تستهدف القضايا الرئيسية والانتخابات والحملات.

قد يكون استخدام القوائم المجمعة لنطاقات المعلومات المضللة أو مواقع الأخبار المزيفة في هذه المرحلة مفيدًا للغاية. جمع الباحثون في ستانفورد قائمة بأكثر من 600 نطاق معروف بإنتاج محتوى زائف في أواخر عام 2018 ، واستهدف بشكل أساسي الولايات المتحدة (Allcott et al, 2018)[29]. يجب مراعاة اثنين من التحذيرات المهمة إذا قررت استخدام مثل هذه القوائم في أي وقت: (1) التضليل والتحرك على الإنترنت بسرعة – ستفقد نطاقات التضليل الجديدة التي حدثت منذ وقت تجميع القائمة عند التحليل ؛ و (2) يجب فحص أي قائمة مستخدمة بشكل صحيح من أجل الدقة المنهجية - قد يؤدي استخدام القوائم المجمعة عشوائيًا الموجودة على الإنترنت إلى الإضرار بسلامة البحث.

تحليل الحسابات والمحتوى

بمجرد قيامك بجمع البيانات و / أو إنشاء شبكة من الحسابات ذات الصلة التي تهتم بها ، تكون أنت وفريقك جاهزين لبدء تحليل الحسابات والمحتوى في مجموعة البيانات. من المحتمل أن يكون هذا هو الجزء من العمل الذي ستقضي فيه معظم الوقت ، خاصة إذا كنت لا تعرف ما تبحث عنه في البداية.

يتم إجراء تحليل المحتوى بشكل أفضل من خلال التأرجح بين الأساليب النوعية والكمية ، وغالبًا ما يتم تكرار العملية عدة مرات للعثور على حسابات محددة أو محتوى مهم. في هذا القسم ، سنقدم لك بعض النصائح والأدوات للمساعدة في توجيه هذا العمل.

28 يُفضل أن يكون جنبًا إلى جنب مع الاستشهاد تقريبيًا لشعبية هذه المصادر. هناك طرق أخرى ، مثل استخراج عناوين URL من أكثر المنشورات التي تمت إعادة تغريدها أو الأكثر إعجابًا في مجموعة بيانات ، وهي أيضًا طرق لتقريب تأثير مجال معين ضمن مجموعة بيانات.

29 يمكنك العثور على هذه الورقة هنا.

أنواع المحتوى

من المفيد فهم أنواع المحتوى الذي قد ترغب في البحث فيه قبل الغوص في مجموعة البيانات.

- **التضليل / المعلومات الزائفة –** يعد المحتوى السياسي الكاذب والمضلل أحد أكثر أشكال المحتوى ضررًا التي تريد مشاهدتها في مجموعة البيانات الخاصة بك. على الرغم من أنك ستسمع محتوى زائف يشار إليه بالعديد من الأسماء ، بما في ذلك الدعاية الحسابية والأخبار المزيفة ، إلا أن هذا النوع من المحتوى يُشار إليه غالبًا باسم التضليل أو المعلومة الزائفة . يكمن الاختلاف التقني بين هذين المصطلحين في القصد من المحتوى الزائف[30]. التعريفات التي سنستخدمها للتمييز بين هاتين الكلمتين مأخوذة من تقرير المعجم عن البيانات والمجتمع (Jack 2017) ، الذي يستكشف مجموعة من الكلمات المفيدة لفهمها واستخدامها عند مناقشة المحتوى الزائف عبر الإنترنت.
- **التضليل –** محتوى كاذب ينتشر بقصد متعمد للخداع. الجهات الفاعلة ذات الدوافع السياسية للدولة أو الجهات ذات الدوافع المالية هي الجهات الأكثر احتمالية لتوزيع المعلومات المضللة عن عمد ، على النحو المحدد بهذه الطريقة.
- **المعلومات الزائفة –** معلومات كاذبة يتم نشرها وتوزيعها دون علم. إذا كان الحساب يروّج لقصة ما دون نية لخداع المستخدمين ، فهذا يعتبر بمثابة معلومات خاطئة.
- **المعلومات الخاطئة –** سلط بعض المؤلفين الضوء أيضًا على ظاهرة "المعلومات الخاطئة" – وهي توزيع معلومات صحيحة أو صادقة في الغالب بقصد الضرر. تعرّف كلير واردل وحسين دراخشان المعلومات الخاطئة على أنها "معلومات قائمة على الواقع ، تُستخدم لإلحاق الأذى بشخص أو منظمة أو دولة" (Wardle and Derakhshan 2017).
- **خطاب الكراهية –** هو لغة تشوه الناس من العرق المستهدف أو الانتماء العرقي أو الجنس أو التوجه الجنسي أو الدين. في كثير من الأحيان ، يحرض خطاب الكراهية الآخرين على التحرش أو التحقير أو حتى المشاركة في العنف ضد الفئة الاجتماعية المستهدفة.

التحليل اللغوي

طريقة واحدة لتحليل المحتوى بطريقة صارمة هي إجراء التحليل اللغوي. يمكن أن يساعدك التحليل اللغوي في فهم اللغات الرئيسية للرسائل ، وموضوعات المحتوى الرئيسية ، والسرد الذي يتم دفعه وتطويره ، والكلمات الرئيسية والمفردات ، وما إذا كان الكلام الذي يحض على الكراهية أو أي محتوى خطير آخر يحدث في المحادثة.

الكلمات الأساسية والمعاجم

من المحتمل ألا تكون كل مشاركة أو تغريدة في مجموعة بيانات معينة ذات صلة بالأسئلة المحددة التي تهمك. لهذا السبب ، قد يكون من المفيد تجميع قائمة بالكلمات الرئيسية ذات الصلة التي من المحتمل أن تحتوي على معلومات ذات صلة باستفسارك. العمل مع خبير في الموضوع – شخص لديه فهم عميق للغة وسياسة المنطقة محل الاهتمام – هو أفضل طريقة لضمان جودة قائمة الكلمات الرئيسية. يُشار أحيانًا إلى قوائم الكلمات الرئيسية هذه باسم "المعاجم" إذا كانت جميعها مرتبطة بموضوع مشترك. على سبيل المثال ، غالبًا ما يتم استخدام معجم خطاب الكراهية بلغات مختلفة مرتبطة بالسياقات السياسية المختلفة لتحليل المحتوى السياسي. حتى جلسة ساعة سريعة مع خبير في الموضوع يمكن أن تقطع شوطًا طويلاً نحو إدخال بعض الدقة في تجميع قوائم الكلمات الرئيسية لضمان الجودة. يمكنك أيضًا الرجوع إلى المؤلفات الأكاديمية ذات الصلة واستخدام قوائم الكلمات الرئيسية التي تم تجميعها مسبقًا إذا كانت ذات جودة عالية وذات صلة بالسياق الخاص بك.

أمثلة للمعاجم

هناك أمثلة وافرة على المعاجم المتوفرة على الإنترنت ومفيدة للتحليل اللغوي وما بعد الاستخراج ذات الصلة. قام فريق النوع الاجتماعي والمرأة والديمقراطية التابع للمعهد الديمقراطي الوطني للشؤون الدولية بتجميع معجم لخطاب الكراهية يتعلق باللغات الإندونيسية والكينية والصربية (Zeiter et al., 2019،).

لدى PeaceTech Lab ، وهي منظمة غير ربحية مكرسة لاستخدام التكنولوجيا لتعزيز السلام في البلدان النامية ، العديد من معجمات خطاب الكراهية المتاحة للجمهور المتاحة مجانًا على موقعها على الإنترنت. هذه المعاجم متعددة اللغات ، والتي يمكن أن تكون مفيدة للغاية لمراقبة المحادثات في مواقع متعددة أو في المناطق ذات التنوع اللغوي العالي.

30 في حين أن هذا هو الحال ، فمن غير الممكن غالبًا معرفة القصد من المحتوى الكاذب الذي ينتشر عبر الإنترنت. لهذا السبب ، من المحتمل أن تسمع هذه الكلمات المستخدمة بالتبادل في بعض الأحيان.

يحتوي Hatebase.org أيضًا على قوائم كلمات رئيسية متعددة اللغات لخطاب الكراهية. استخدم الباحثان رويا باكزاد ونيلوفر صالحي قائمة تم تجميعها من هذا الموقع لدراستهما في الدعاية الحسابية التي تستهدف الأمريكيين المسلمين في منتصف الانتخابات النصفية 2018 (2019). تم استخدام قائمة مماثلة للتحليل الكمي لرهاب الإسلام على Gab في الفترة التي سبقت الانتخابات نفسها (وولي ، باكزاد وموناكو ، 2019).

التحليل اللغوي النوعي

بمجرد الحصول على مجموعة من المنشورات ذات الصلة المستخرجة ، يمكنك الغوص في التحليل اللغوي. الطرق النوعية هي تلك التي تحلل موضوعات المحتوى والرسائل للمشاركات في مجموعة البيانات الخاصة بك. يكون هذا النوع من العمل أكثر فاعلية عندما يقوم به الأفراد ويقوم به خبير أو فريق خبراء على دراية بالسياق السياسي واللغوي للمنطقة المعنية.

سواء باستخدام الأساليب النوعية أو الأساليب الكمية أو كليهما أثناء التحليل الخاص بك ، فمن المهم للغاية أن تكون أساليبك متسقة ومنهجية. يعد استخدام نفس الأساليب لتحليل كل منشور أو جزء من مجموعة بياناتك أفضل طريقة للوقاية من إدخال التحيز في نتائجك.

التحليل السردي

يعد التحليل السردي طريقة مكثفة لتحليل المحتوى اللغوي في مجموعة بيانات ، ولكن يمكن أن يكون مثيرًا للاهتمام. يمكن أن يلقي التحليل السردي الذي يحلل كيف تشير بعض المنافذ إلى موضوع معين ، خاصة بمرور الوقت ، الضوء على تأطير وتأثير المنافذ على الرأي العام.

الترميز النوعي

يُشار إلى طريقة نوعية أخرى مفيدة لفهم أنواع الرسائل التي تحدث في مجموعة بيانات معينة باسم الترميز النوعي. ويتكون من فريق من الخبراء يعين بشكل مستقل واحدة من مجموعة من الفئات المحددة مسبقًا (أو "الرموز") للمشاركات واحدًا تلو الآخر. بعد تخصيص الفئات لكل تغريدة ، يمكن إجراء التحليل الكمي.

على سبيل المثال ، في دراسة افتراضية حول انتخابات رئاسية بين مرشحين - المرشح أ والمرشح ب - في دولة قمر[31] الخيالية ، يمكننا أن نتخيل خمس فئات ممكنة للرسائل التي تم جمعها حول الانتخابات:

1. مؤيد المرشح أ
2. مؤيد المرشح ب
3. معارض المرشح أ
4. معارض المرشح ب
5. المحايد

بعد أن قام فريق من الخبراء بتشفير جميع المنشورات في مجموعة البيانات ، هناك العديد من التحليلات الكمية التي يمكننا القيام بها كخطوة تالية:

- **توزيع المشاركات في كل فئة -** يمكننا تحليل كمية المشاركات التي تحدث في كل فئة لفحص ما إذا كان هناك دعم أكبر أو معارضة أكبر عبر الإنترنت للمرشح أ أو المرشح ب .
- **استخدام الكلمات الرئيسية حسب الفئة -** يمكننا أيضًا فحص المشاركات من كل فئة بحثًا عن الكلمات الرئيسية ذات الصلة ، لمعرفة ما إذا كان مؤيدو أو معارضو أي من المرشحين يستخدمون كلمات معينة.
- **محتوى البوت في كل فئة -** هناك احتمال آخر وهو مراجعة المشاركات من كل فئة لمحتوى البوت لمعرفة ما إذا كان أي من المرشحين يحصل على دعم أكبر أو معارضة من وكلاء مؤتمين عبر الإنترنت.

هذه ليست سوى أمثلة قليلة من التحليلات التي يمكن إجراؤها بعد إجراء التشفير النوعي عالي الجودة.

31 قمر بلد خيالي يظهر في مسلسل الدراما السياسية الأمريكية . الجناح الغربي.

التحليل اللغوي الكمي

يُعرف التحليل اللغوي الكمي أيضًا بمعالجة اللغة الطبيعية (NLP) أو اللغويات الحاسوبية. تنبع الأساليب المستخدمة في تحليل اللغويات الكمية من فكرة أنه يمكننا اكتساب رؤى معينة حول أنواع وتكرار رسائل معينة عندما ننظر إلى اللغة من وجهة نظر إحصائية. بعبارة أدق ، إذا قمنا بتجميع الكلمات أو مجموعات الكلمات من المنشورات ذات الصلة في مجموعة بياناتنا ، يمكننا فحص أنواع المواضيع التي تظهر. يقدم لك هذا القسم الفرعي بعض أساسيات استخراج عدد الكلمات من مجموعة من البيانات لفحص الموضوعات الشائعة عبر الإنترنت.

بينما يمكن تنفيذ التقنيات في هذا القسم بأي لغة برمجة ، تجدر الإشارة إلى أن Python و R يأتيان مع عدة حزم تبسط العملية. تحتوي هذه الحزم ، مثل مجموعة أدوات اللغة الطبيعية (NLTK) من Python أو حزمة R's TM (وهي اختصار لمصطلح تنقيب النصوص) على وثائق شاملة في شكل كتب وأدلة عبر الإنترنت وبرامج تعليمية يمكن أن تعلمك الأساسيات في غضون بضع ساعات. نوصي بشدة تخصيص بعض الوقت للتعرف على إحدى هذه الحزم ، خاصة إذا كانت لديك معرفة مسبقة بحزم Python أو R فهي توعد عليك وعلى فريقك الوقت في المستقبل!

الن-غرام (N-gram)

هي اللبنة الأساسية لجميع التحليلات اللغوية الكمية لمحتوى الوسائط الاجتماعية تقريبًا. وهي عبارة عن سلسلة من الكلمات ذات الطول n. هناك ثلاثة أنواع خاصة بها تعتبر الأكثر شيوعًا للتحليل اللغوي الكمي للنصوص:

- **Unigrams –** يمكن اعتبار كلمة واحدة بمثابة تسلسل لكلمة واحدة. يُعرف هذا النوع من n-gram باسم unigram.
- **Bigrams -** زوج من الكلمات التي تحدث بجانب بعضها البعض في جملة تشكل bigram. وبعبارة أخرى ، فإن bigram هو أي تسلسل لكلمتين في النص.
- **Trigrams -** حتى الآن ، ربما تكون قد خمنت أن Trigram هو أي ثلاث كلمات في جملة. كل تسلسل من 3 كلمات في جملة واحدة أو نص يشكل مثلثًا منفصلًا.

من الشائع العمل مع هذه الأنواع الثلاثة من n-grams. ويرجع ذلك إلى حد كبير إلى حقيقة أن استخدام قيم أكبر لـ n هو "تكلفة حسابية" - من المحتمل أن تبطئ أداء جهاز الكمبيوتر الخاص بك ولا تحصل على الكثير من المال مقابل ذلك. بالنسبة إلى تسلسل الكلمات الأكبر من 3 ، فمن الشائع الإشارة ببساطة إلى هذه الأرقام في العدد متبوعًا بكلمة "-gram". على سبيل المثال ، 4-grams ستكون عبارة عن سلسلة من 4 كلمات ، 5- grams ستكون سلسلة من 5 كلمات ، إلخ.

لفكرة الرئيسية وراء n-grams هي أنه ليس لديك فقط عدد الكلمات ، ولكن لديك أيضًا كمية صغيرة من السياق الذي تحدث فيه الكلمة. مع العلم أن لديك 16 مرة من كلمة "ملك " تخبرك أن لديك عددًا كبيرًا من الكلمات الموضوع على يدك ، ولكن ليس أكثر من ذلك. إذا قمت بتوسيع تحليلك إلى 4- grams ورأيت أن لديك 15 مرة من "يسقط الملك" وحدث واحد "يعيش الملك" ، يمكنك القول على وجه اليقين أن معظم المشاركات ذات الصلة بـ "الملك" في مجموعة البيانات الخاصة بك ليست رسائل دعم.

بمجرد أن يكون فريقك مرتاحًا لتقنية استخراج n-grams واحتساب ترددات n-gram داخل النص ، يمكنك تطبيق هذه التقنية على أجزاء مختلفة من مجموعة البيانات الخاصة بك. على سبيل المثال ، قد تكون مقارنة n-grams بين الحسابات أو الصفحات التي تدعم مرشحين / أحزابًا سياسية مختلفة مفيدة. إن مقارنة ترددات n-gram من وسائل الإعلام المختلفة التي تنتج مقالات ذات صلة بالانتخابات نفسها قد تعطي نظرة ثاقبة على التركيز الرئيسي لكل منفذ. قد توضح مقارنات تكرار N-gram بين تاريخ النشر لحسابين أو صفحتين منفصلتين موضوعات المراسلة المفضلة لديهم. هذه ليست سوى أمثلة قليلة للإمكانيات التي يمكن أن تساعد في تحسين بحثك.

التقنيات اللغوية الكمية الأخرى

تجميع ترددات n-gram هي تقنية يمكن تطبيقها كخطوة أولى لتقنيات البرمجة اللغوية العصبية الأخرى التي يمكن أن تساعد في تحليل مجموعة بيانات بتفصيل أكبر. على سبيل المثال ، بعد استخراج ترددات الكلمات أو ترددات n-gram ، يمكن استخدام تقنية إحصائية تعرف باسم تردد المستند العكسي للتردد (tf-idf) لمقارنة موضوعات الرسائل لمستندات مختلفة بالنسبة لبعضها البعض. يمكن أن تكون هذه "المستندات" عبارة عن مجموعات من المقالات من مختلف المنافذ الإخبارية المؤدية إلى الانتخابات أو نشر التواريخ لحسابات مختلفة ذات أهمية ، على سبيل المثال. Tf-idf هو طريقة أساسية لتحديد السمات الفريدة التي تميز مستندًا عن آخر. يمكن استخدام هذه التقنية لتحليل الحسابات التي غالبًا ما تروج لطرف معين ، أو ما الذي يجعل رسائل مجتمع ما مميزة عن جميع الآخرين الذين تفحصهم.

يمكن العثور على دليل سريع حول كيفية استخدام tf-idf للتحليل اللغوي[32] وتجميع المحتوى في Mining the Social Web (Russell and Klassen 2018). هذه المناقشة تستحق التشاور - فهي تحتوي على رمز مثال لحزمة Python ويسهل متابعتها بشكل خاص.

كشف خطاب الكراهية الآلي

لتحليل خطاب الكراهية في مجموعة بيانات ، ستضطر على الأرجح أولاً إلى تجميع قائمة بمصطلحات الخطاب ذات الصلة بالمنطقة التي تفحصها. يجب مراجعة هذه المصطلحات مع خبير في الموضوع بالتفصيل وفي كل لغة من المحتمل استخدامها في المنطقة المعنية ، والمنهجية الموصوفة في تقرير المعهد الديمقراطي الوطني "Tweets That Chill". (Zeiter et al., 2019)

أحد أساليب الكشف التلقائية عن خطاب الكراهية بدون تجميع الكلمات الرئيسية هو استخدام واجهة برمجة تطبيقات Google Jigsaw Perspectives - وهي أداة مفتوحة المصدر تم تقديمها في عام 2017. وهي فقط تدعم قراءة اللغة الانجليزية في التعليقات[33] حاليا. تأخذ الأداة سلسلة من النص كمدخلات وتخرج درجة الكراهية لهذا النص. وكلما ارتفعت الدرجة ، أصبح الكلام أكثر "كراهية" من خلال نماذج التعلم الآلي في واجهة برمجة تطبيقات Perspectives. يمكن العثور على مثال لاستخدام واجهة برمجة التطبيقات هذه في مجموعة بيانات Gab الخاصة بـ Pushshift.io. بالإضافة إلى احتواء المنشورات من Gab، Pushshift.io المنقحة من خلال واجهة برمجة تطبيقات Perspectives وتسجيل درجة الكراهية التي تلقتها.

تعد واجهة برمجة التطبيقات Perspectives حاليًا واحدة من الأدوات الوحيدة التي تحدد درجات خطاب الكراهية المتاحة للجمهور. يستغرق الأمر وقتًا وجهدًا كبيرًا لالتقاط الفروق الدقيقة في السياق والقصد عند التطرق للغات البشرية ، ولهذا السبب ، لا تزال أدوات تحليل المشاعر واكتشاف الكراهية في مرحلة الطفولة. علاوة على الصعوبة العامة الكامنة في المشكلة ، فإن الخصوصية المتنوعة لسياق كل لغة والسياق الاجتماعي الذي يحدث فيه خطاب الكراهية يضاعف من صعوبة إنتاج أدوات كشف خطاب الكراهية الآلية الصارمة والموثوقة. ولهذه الأسباب ، لا يزال العمل جنبًا إلى جنب مع خبراء الموضوع لتجميع الكلمات الرئيسية والمفردات ذات الصلة أفضل طريقة لتحليل خطاب الكراهية في مجموعة بيانات وسائل الإعلام الاجتماعية.

الروبوتات

لقد كتب الكثير عن تأثير الروبوتات على وسائل التواصل الاجتماعي في السنوات القليلة الماضية. قام مشروع الدعاية الحاسوبية (ComProp) في جامعة واشنطن ومعهد أكسفورد للإنترنت (OII) بعمل رائد في برامج الروبوت ، واستكشف تأثيرها في نشر وتعزيز الدعاية الحسابية في البلدان حول العالم.

الروبوتات ، ببساطة ، هي برامج كمبيوتر تتحكم في الملفات الشخصية على مواقع التواصل الاجتماعي ، والتي غالبًا ما تتظاهر بأنها أشخاص حقيقيين وتتفاعل مع البشر الآخرين عبر الإنترنت. في حالة وسائل التواصل الاجتماعي ، يتم التحكم في معظم برامج الروبوت من خلال برامج الكمبيوتر التي تتحكم في الحسابات من خلال واجهة برمجة التطبيقات[34]. غالبًا ما تحتوي أبسط هذه الروبوتات على علامات منبهة تجعلها تلقائية - تغرد في نفس الدقيقة من كل ساعة أو لا تحتوي على صورة ملف شخصي ، على سبيل المثال. غالبًا ما تُستخدم هذه البيانات في خوارزميات التعلم الآلي للأدوات التي تفرز الروبوتات عن البشر عبر الإنترنت.

32 سيقدم لك هذا النص أيضًا بعض التقنيات ذات الصلة لتحسين تحليلك ، مثل إزالة كلمات التوقف وإيقافها.

33 استفادت اللغة الإنجليزية بشكل غير متناسب من اهتمام اللغويين ، سواء الحسابي أو غير الحسابي ، طوال القرنين العشرين والحادي والعشرين. غالبًا ما يُشار إلى اللغات التي لم تستفد من الاهتمام والبحث الكبيرين على أنها لغات منخفضة الموارد في اللغويات والبرمجة اللغوية العصبية. لا يمكن المبالغة في أهمية إجراء البحوث على هذه اللغات. إذا أجريت أنت أو فريقك خطابًا يحض على الكراهية أو بحثًا لغويًا كجزء من جهود مراقبة وسائل الإعلام الخاصة بك ، فقد يكون من المفيد التواصل مع اللغويين المحترفين في الصناعة أو الأوساط الأكاديمية لمعرفة ما إذا كان البحث الخاص بك يمكن أن يساعد في تقديم رؤى قيمة في المجال.

34 في قسم جمع البيانات أعلاه ، اكتشفنا الاختلافات بين واجهات برمجة التطبيقات وكشط الويب - وهذا الاختلاف مفيد أيضًا لفهم كيفية عمل بعض الروبوتات على وسائل التواصل الاجتماعي وعلى الويب بشكل عام دون استخدام واجهات برمجة التطبيقات. يمكن برمجة برامج الروبوت للتفاعل مع صفحات الويب اليومية - في الواقع ، هذا أمر نافه تمامًا للمبرمجين للقيام به. على سبيل المثال ، يمكن أن تنتقل برامج التتبّع من موقع إلى آخر وتحليل محتوى صفحات الويب - غالبًا ما تسمى هذه الأنواع من برامج التتبّع ببرامج الزحف أو العناكب ، وفي الواقع ، هذه هي الطريقة التي تجمع بها Google البيانات على صفحات الويب لوضعها في محركات البحث. كما أن ليس كل برامج التتبّع روبوتات على وسائل التواصل الاجتماعي ، وليس جميعها سيئة ، وفي الواقع بعضها ضروري لاستخدام الانترنت اليومي.

أدوات الكشف عن الروبوتات

على الرغم من عدم وجود طريقة للتأكيد ما إذا كان الحساب روبوتًا أم لا ، إلا أن هناك خيارات مفيدة متاحة للباحث النهم للكشف عن الروبوتات عبر الإنترنت. أحد أفضل الخيارات المتاحة حاليًا هو Botometer ، وهي أداة تستخدم التعلم الآلي لتعيين درجة احتمالية الروبوت لحسابات الإدخال. لدى Botometer تاريخ طويل من البحث والتطوير الأكاديمي في جامعة إنديانا وهو مجاني للاستخدام. يوجد أيضًا عدد من المصنفات الأخرى (انظر الجدول أدناه).

أدوات الكشف عن الروبوتات

اسم الأداة	هل يمكن تنفيذه من خلال كود تصنيف دفعات الحسابات؟	المنصة[35]	ملحقات / موقع متاح؟
Botometer	نعم (Python)	Twitter	يمكن التحقق من الحسابات الفردية على الموقع.
Tweetbotornot	نعم	Twitter	متاح فقط تنفيذ التعليمات البرمجية.
Botcheck.me	لا	Twitter	ملحقات Chrome و Firefox ، آخر تحديث في 2019.
Botsentinel	لا	Twitter	تطبيقاتAndroid و ملحقات Chrome و Firefox
Pegabot	لا	Twitter	الموقع

35 تم تطوير أدوات الكشف عن الروبوتات مفتوحة المصدر فقط على Twitter - ويرجع ذلك إلى حد كبير إلى أن واجهة برمجة تطبيقات Twitter تقدم مجموعة غنية من المعلومات حول المستخدمين (البيانات الوصفية العامة) ، والتي تعد مفيدة كميزات في إنشاء نماذج التعلم الآلي التي تصنف الحسابات على أنها برامج الروبوت أو الأشخاص. يعتمد تحليل نشاط برنامج التتبّع على الأنظمة الأساسية الأخرى إلى حد كبير على التحليل اليدوي والتحقيق ، مثل اكتشاف نشاط خارق للإنسان (100 مشاركة / دقيقة ، ونشر الرسائل على فترات منتظمة ، وما إلى ذلك). على سبيل المثال ، في دراسة حول رهاب الاسلام التي سبقت الانتخابات النصفية في أمريكا عام 2018 على منصة وسائل التواصل الاجتماعي Gab ، اكتشفنا وجود روبوت مضلل من خلال تحليل المنشورات المتطابقة التي تحدث في فترات قصيرة من مستخدم واحد.

الـــخـــاتـــمة

في نهاية المطاف ، يجب أن تساعد كل هذه التقنيات والأدوات في تشكيل نهج المستخدم في جمع البيانات وتحليلها. يجب على المستخدمين أيضًا التفكير في بناء سير عمل باستخدام هذه التقنيات ، ونظام أرشفة وفرز المعلومات التي تم جمعها ، والإبلاغ المحتمل عنها إلى المنصات أو الوكالات الوطنية الأخرى التي يمكنها التصرف بناءً على البيانات التي تم جمعها. اعتمادًا على الموضوع ، يجب عليهم التفكير في عدد من الأدوات والتقنيات المختلفة الموضحة أعلاه. سواء كان البحث عن معلومات مضللة أثناء الانتخابات أو خطاب الكراهية والدعاية الحسابية في الخطاب السياسي التقليدي عبر الإنترنت ، يمكن تطبيق هذه الموارد بطرق مختلفة. يجب على الفرق والباحثين أيضًا التفكير في الموارد التي لديهم - بشرية ومالية وتقنية - أثناء إنشاء مشروعهم.

العديد من الأدوات المذكورة مفتوحة المصدر ، ولكن البعض الآخر باهظ الثمن وغالبًا ما يكون معقدًا للتطبيق دون معرفة متقدمة بالأدوات والأساليب المستخدمة في تطويرها. غالبًا ما يجدر التفكير في حلول أبسط يمكن تطبيقها من قبل الباحثين الأقل خبرة ، أو العمل بشكل تعاوني عبر فرق ذات مجموعات مهارات مختلفة لاكتشاف رؤى مختلفة في نفس مجموعة البيانات. خذ بعين الاعتبار الشراكة مع باحثين محليين وخبراء تقنيين دوليين للحصول على أنواع جديدة من الأفكار ، والعمل على بناء طرق للتعاون عبر الإنترنت عبر المجتمعات والبلدان والمناطق.

إحدى الطرق للعثور على شركاء للتعاون معهم ، في سياقك المحلي والدولي ، وللتواصل مع شركات وسائل التواصل الاجتماعي لإعداد التقارير والبحث ، من خلال منصة Design 4 Democracy Coalition. منصة D4D هي مجموعة من المنظمات غير الحكومية الدولية بما في ذلك المعهد الديمقراطي الوطني للشؤون الدولية والمعهد الجمهوري الدولي والمؤسسة الدولية للأنظمة الانتخابية والمعهد الدولي للديمقراطية ومساعدة الانتخابات ومنظمات المجتمع المدني المحلية في جميع أنحاء العالم التي تتعامل مع شركات التكنولوجيا مثل Facebook و Microsoft و Twitter لتشجيعهم في تصميم أنظمتهم ، وإدارة المحتوى وسياسات للمبادئ الديمقراطية. أصبحت الأبحاث والمراقبة عبر الإنترنت مكونين حاسمين للانتخابات والديمقراطيات في جميع أنحاء العالم ، وقد تم كتابة هذا الدليل لإعطاء المجموعات العاملة في هذا المجال الدعم وذلك من خلال السياسات والأنظمة الفنية الأدوات والأساليب والقدرات للقيام بهذا العمل والمساعدة في إعلام المجتمع بشكل أفضل.

انظر أدناه للحصول على مزيد من المراجع والأدوات مفتوحة المصدر وأمثلة على التعليمات البرمجية وشرح تفصيلي لاستخدام واجهة برمجة تطبيقات Twitter.

الملحق 1: مثال على رمز واجهة برمجة التطبيقات - جمع البيانات من واجهات برمجة تطبيقات البحث والتدفق من Twitter باستخدام حزمة إعادة التغريد

في هذا القسم ، سنقدم لك بإيجاز بعض الرموز التي يمكنك استخدامها لجمع بيانات Twitter التاريخية والحيوية. من خلال بضعة أسطر من التعليمات البرمجية ، يمكنك بسهولة جمع بيانات Twitter ذات الصلة بالسياق الانتخابي. بعد جمع البيانات المستهدفة ، يمكنك تصدير البيانات المطلوبة إلى CSV ، أو متابعة تحليلها في Excel أو جداول بيانات Google ، أو تمرير هذه البيانات إلى عالم بيانات مخصص في فريقك للبحث عن رؤى أعمق. إن القدرة على جمع البيانات بنفسك في الوقت الفعلي ، حتى لو لم يكن لديك عالم بيانات مخصص في فريقك في وقت التجميع ، أمر مفيد للغاية. وعلى العكس ، غالبًا ما تصبح البيانات أكثر قيمة بمرور الوقت لأنها تلتقط معلومات مضللة وبرامج الروبوت والممثلين الشائنين الذين قد تتم إزالة أفعالهم ومشاركاتهم لاحقًا من النظام الأساسي لانتهاك شروط الخدمة.

الخطوة 1: قم بتنزيل R و RStudio و Rtweet

في هذا الدليل ، سنعلمك سحب بعض البيانات الأساسية من واجهة برمجة تطبيقات Twitter باستخدام لغة البرمجة R. هناك حزمة محددة في R ، تسمى rtweet ، تجعل سحب البيانات من Twitter أمرًا سهلاً للغاية.

سيتعين عليك تنزيل بعض العناصر وتثبيتها لبدء التشغيل.

- Rstudio: https://www.rstudio.com/products/rstudio/download
- R: https://www.r-project.org

الخطوة 2: التقدم بطلب للحصول على تطبيق Twitter ، واسترداد مفاتيح التطبيق وإنشاء الرمز المميز الخاص بك

من أجل جمع البيانات من Twitter ، يمكنك استخدام تطبيق - هذه وسيلة آمنة لاستخدام حسابك للتفاعل مع النظام الأساسي من خلال التعليمات البرمجية. في السنوات الماضية ، كانت تطبيقات Twitter مجانية ولم تكن هناك حاجة للموافقة - في الوقت الحاضر ، عليك التقدم بطلب للحصول على تطبيق ، والذي يمكنك القيام به هنا.

بمجرد استرداد التطبيق الخاص بك ، ستحتاج إلى تسجيل الدخول إلى Twitter على متصفح والتوجه إلى https://apps.twitter.com. يمكنك استرداد مفاتيح المستهلك والوصول الخاصة بتطبيقك هنا - سيكون هناك أربعة مفاتيح إجمالية. هذه المفاتيح هي ببساطة سلاسل نصية سيستخدمها تطبيقك للتحقق من أنه يطلب بالفعل بيانات نيابة عن شخص - أنت في هذه الحالة. هذا الاستخدام لمفاتيح المصادقة هو عملية تعرف باسم oauth ، والتي تعني المصادقة المفتوحة. تم تطويرها كوسيلة للسماح للتطبيقات بتنفيذ الإجراءات عبر الإنترنت نيابة عن المستخدمين دون نقل كلمة المرور الخاصة بهم واسم المستخدم لكل إجراء.

بمجرد الحصول على مفاتيحك واسم التطبيق الخاص بك ، يمكنك استخدام الرمز الذي يتبع في لقطة الشاشة أدناه لإنشاء "رمز" التفويض الخاص بك. بمجرد إنشاء هذا الرمز المميز ، تكون مستعدًا لبدء استخدام Twitter API.

```
library(rtweet)
app<-"<app name here>"
consumer_key<-"<consumer key here>"
consumer_key_secret<-"<consumer key secret here>"
access_token<-"<access token here>"
access_token_secret<-"<access token secret here>"
create_token(app,consumer_key, consumer_key_secret, access_token, access_token_secret)
```

أسطر R Code أعلاه هي خطوط أولية يمكنك استخدامها لمصادقة تطبيق Twitter الخاص بك والاتصال بـ Twitter API. بمجرد التقدم بطلب للوصول إلى تطبيق Twitter والموافقة عليه ، يمكنك استرداد المستهلك والوصول إلى الرموز المميزة من Twitter. ستستخدم هذه الرموز المميزة لمصادقة تطبيقك (كما هو موضح في الرمز أعلاه). بمجرد مصادقة تطبيقك (على سبيل المثال ، يعرف Twitter أن التطبيق يسحب البيانات لك وليس لشخص آخر) ، فأنت على استعداد لبدء التفاعل مع البيانات وسحبها من Twitter API.

يشرح الأستاذ في جامعة ميسوري ومطور rtweet الدكتور Mike Kearney أيضًا عملية المصادقة خطوة بخطوة على الموقع الرسمي لـ rtweet هنا.

جمع البيانات القديمة:

باستخدام rtweet ، يمكنك بسهولة جمع التغريدات باستخدام واحدة أو أكثر من علامات التصنيف ذات الأهمية من بحث Twitter API. هذه الواجهة قديمة ، مما يعني أنك ستجمع بيانات بتاريخ الأيام 9-7 الماضية التي تتوافق مع استعلامك. الاستفسارات هي ببساطة معايير تهمك في التغريدات التي تجمعها. يمكن أن تحتوي الاستعلامات على علامات الوسم، أو كلمات رئيسية ، أو أسماء حسابات ، أو عناوين URL ، أو مزيج منها ، أو جميعها. في هذا القسم ، سنركز على علامات التصنيف ، ولكن العملية والصيغة هي نفسها تمامًا للكيانات الأخرى.

الوظيفة الرئيسية التي ستستخدمها للبحث عن التغريدات القديمة تسمى search_tweets ().

```
my_data<-search_tweets("#ExampleHashtag", n=50000, retryonratelimit=TRUE)
```

يستعلم هذا السطر من كود R عن واجهة برمجة تطبيقات بحث Twitter لما يصل إلى 50000 من التغريدات التي تستخدم #Example Hashtag من آخر 9-7 أيام ، ويحفظ النتائج في إطار بيانات يسمى "my_data". إذا استخدمت علامة الوسم لأقل من 50000 مرة في هذا الإطار الزمني ، فسيتم إرجاع عدد أقل من التغريدات. يمكن أن يؤدي تغيير قيمة الإدخال لـ "n" إلى زيادة أو تقليل الحد الأقصى للتغريدات التي ستستردها.

```
multiple_hashtags<-search_tweets("#ExampleHashtag1 OR #ExampleHashtag2", n=50000, retryonratelimit=TRUE)
```

يشبه هذا السطر من كود R الاستعلام أعلاه ، ولكنه يعرض ما يصل إلى 50000 تغريدات تحتوي على # ExampleHashtag1 أو # ExampleHashtag2. يمكن استخدام هذه البنية للاستعلام عن واجهة برمجة تطبيقات بحث Twitter عن علامات التصنيف المتعددة. في رمز المثال هذا ، يتم حفظ النتائج في dataframe يسمى "multi_hashtags". يتم فصل علامات الوسم المتعددة بما يسمى "عوامل التشغيل المنطقية": وهي ببساطة "AND" أو "OR". استخدم "AND" إذا كنت تريد فقط التغريدات التي تحتوي على علامات الوسم ؛ ستُرجع "أو" التغريدات التي تحتوي على أيٍ منهما.

تدفق بيانات Twitter في الوقت الفعلي:

لديك أيضًا خيار دفق بيانات Twitter في الوقت الفعلي. يتطلب هذا الاستعلام تحديد المدة الزمنية التي ترغب في دفق التغريدات فيها في ثوانٍ ، بالإضافة إلى الكيانات التي ترغب في دفقها (علامات التصنيف ، عناوين URL ، الكلمات الرئيسية ، @ الإشارات ، وما إلى ذلك).

تختلف بنية دفق التغريدات في rtweet قليلاً عن تلك المستخدمة في الاستعلام عن واجهة برمجة تطبيقات البحث. عند بث التغريدات ، ستحتاج إلى استخدام وظيفة stream_tweets (). سيتم فصل عناصر الاستعلام المتعددة بفواصل.

```
my_streamed_data<-stream_tweets(q="#ExampleHashtag1",timeout=60)
```

```
my_streamed_data_w_multiple_hashtags<-stream_tweets(q="#ExampleHashtag1,#ExampleHashtag2",timeout=60)
```

توضح الأشكال أعلاه كيفية استخدام الدالة stream_tweets () في rtweet لجمع التغريدات في الوقت الفعلي. يجمع الشكل العلوي أي تغريدات باستخدام # ExampleHashtag1 خلال نافذة بث 60 ثانية. الرقم السفلي يفعل الشيء نفسه ، ولكنه يجمع أيضًا التغريدات التي تحتوي على # ExampleHashtag2.

كتابة النتائج إلى ملف إخراج CSV:

تنسيق ملف شائع لتحليل البيانات هو ملف CSV ، والذي يعني "قيم مفصولة بفواصل". في ملف CSV ، يمثل كل سطر صفًا واحدًا من جدول البيانات[36] ، ويمثل كل كيان بين الفواصل خلية - أو بشكل أدق ، قيمة[37] داخل الخلية.

ملف CSV هو في الأساس جدول بيانات بتنسيق قابل للقراءة آليًا. بمجرد حصولك على ملف CSV للتغريدات التي قمت بجمعها ، يمكنك نقلها إلى عالم بيانات متخصص لإجراء تحليل أعمق ، أو تحميلها في معالج جداول البيانات مثل Microsoft Excel أو جداول بيانات Google لإجراء بعض التحليلات بنفسك.

36 يُشار أحيانًا إلى هذه الصفوف ببعض المصطلحات الأخرى من قبل علماء البيانات: التسجيل والمثال والملاحظة هي مصطلحات أخرى من المحتمل أن تسمعها في هذا السياق ، وكلها مرادفة لـ "الصف" عند الإشارة إلى ملف csv.

37 يمكن أحيانًا الإشارة إلى القيم الموجودة في ملف csv أو جدول البيانات على أنها حقول.

```
write_as_csv(my_data,"my_data_as_a_csv_file.csv")
```

يستخدم هذا السطر من كود R حزمة rtweet لكتابة إطار بيانات من التغريدات يسمى "my_data" إلى ملف CSV يسمى "my_data_as_a_csv_file.csv".
بعد تصدير تغريداتك إلى ملف CSV ، يمكن مشاركة بياناتك أو استيرادها بسهولة إلى Microsoft Excel أو جداول بيانات Google.

الخلاصة النهائية

إن القدرة على سحب البيانات من Twitter وكتابتها في ملف CSV هي مهارة لا تقدر بثمن. وحتى بدون خبرة البرمجة ، يمكن لأي شخص تعلم ما تم وصفه هنا في أقل من ساعتين ، وربما بشكل أسرع. بمجرد أن تكون لديك القدرة على استرداد وتخزين ملفات CSV لبيانات Twitter ذات الصلة ، ستتمكن من وضع نفسك وفريقك في وضع يسمح لهم بتحليل البيانات القيمة الآن وفي المستقبل.

الملحق الثاني: أدوات OSINT

الذكاء المفتوح المصدر ، الذي يشار إليه عادة باسم OSINT ، هو فن التحقيق في سؤال باستخدام المعلومات والبيانات المتاحة للجمهور فقط (أو "المصدر المفتوح"). في سياق المعلومات المضللة ومراقبة وسائل التواصل الاجتماعي ، يمكن لـ OSINT غالبًا تقديم تفاصيل إضافية حول الجهات الشائنة أو المشبوهة عبر الإنترنت ، بما في ذلك الحسابات المزيفة أو مواقع المعلومات المضللة. فيما يلي قائمة ببعض الموارد القيمة لتعلم تقنيات OSINT.

- كتاب وموقع مايكل بازل:
  - https://inteltechniques.com
  - Open-Source Intelligence Techniques
- مجموعة أدوات التحقيق عبر الإنترنت من Bellingcat:
  https://docs.google.com/document/d/1BfLPJpRtyq4RFtHJoNpvWQjmGnyVkfE2HYoICKOGguA/edit
- القائمة العامة للاختصارات و أدوات OSINT للصحفي كريغ سيلفرمان:
  https://docs.google.com/document/d/1ZJbIUk5L8fe3VKK9CLVNMj9qOFdXG-RhQT6pyEgsS4I/edit
- Comprop Navigator ، الذي نشره مشروع الدعاية الحاسوبية لمعهد أكسفورد للإنترنت ، يجمع الأساليب وأدوات OSINT المتعلقة بالمعلومات وأبحاث أخرى عبر الإنترنت.
  https://navigator.oii.ox.ac.uk/
- المسودة الأولى لدليل تجميع الأخبار والمراقبة على مواقع التواصل الاجتماعي
  https://firstdraftnews.org/wp-content/uploads/2019/10/Newsgathering_and_Monitoring_Digital_AW3.pdf?x36710
- محاربة التضليل على الإنترنت: قاعدة بيانات لأدوات الويب ، تستضيفها مؤسسة Rand.
  https://www.rand.org/research/projects/truth-decay/fighting-disinformation.html
- دليل التحقق من التضليل والتلاعب في وسائل الإعلام
  https://datajournalism.com/read/handbook/verification-3/

الـمـراجـع

Allcott, H., Gentzkow, M., & Yu, C. (2018). Trends in the Diffusion of Misinformation on Social Media. https://web.stanford.edu/~gentzkow/research/fake-news-trends.pdf

Borgatti, S., Everett, G., & Johnson, J. (2013). Analyzing Social Networks.

Democracy Reporting International. (October, 2019) *Guide for Civil Society on Monitoring Social Media During Elections.* https://democracy-reporting.org/wp-content/uploads/2019/10/social-media-DEF.pdf

Jack, C. (2017). *Lexicon of Lies.* Data & Society. https://datasociety.net/output/lexicon-of-lies/

Monaco, N. (2019). *Welcome to the Party: A Data Analysis of Chinese Information Operations.* Retrieved from https://medium.com/digintel/welcome-to-the-party-a-data-analysis-of-chinese-information-operations-6d48ee186939

National Democratic Institute. (May, 2019) *Disinformation and Electoral Integrity: A Guidance Document for NDI Elections Programs.* https://www.ndi.org/publications/disinformation-and-electoral-integrity-guidance-document-ndi-elections-programs

National Democratic Institute. (December, 2018). *Supporting Information Integrity and Civil Political Discourse.* https://www.ndi.org/publications/supporting-information-integrity-and-civil-political-discourse

Pakzad, R., & Salehi, Ni. (2019). Anti-Muslim Americans: Computational Propaganda in the United States. Institute from the Future. Retrieved from http://www.iftf.org/fileadmin/user_upload/downloads/ourwork/IFTF_Anti-Muslim_comp.prop_W_05.07.19.pdf

Russell, M., & Klassen, Mikhail. (2018). *Mining the social web.* Sebastopol, CA: O'Reilly Media.

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making.* Council of Europe. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Woolley, S., Pakzad, R., & Monaco, N. (2019). *Incubating Hate: Islamophobia and Gab.* German Marshall Fund. http://www.gmfus.org/sites/default/files/publications/pdf/Incubating%20Hate%20-%20Islamophobia%20and%20Gab.pdf

Woolley, S., & Howard, P. (2017). *Computational Propaganda Worldwide: Executive Summary.* Working Paper. 2017.11. Oxford, UK: Project on Computational Propaganda. http://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf

Zeiter, K., Pepera, S., Middlehurst, M., Ruths, D. (2019). *Tweets That Chill: Analyzing Online Violence Against Women in Politics.* National Democratic Institute. https://www.ndi.org/tweets-that-chill

NATIONAL
DEMOCRATIC
INSTITUTE

NDI.ORG