WHAT ARE THE LOCALS SAYING RELATED TO THE MIGRATION?

RESULTS ANALYSIS

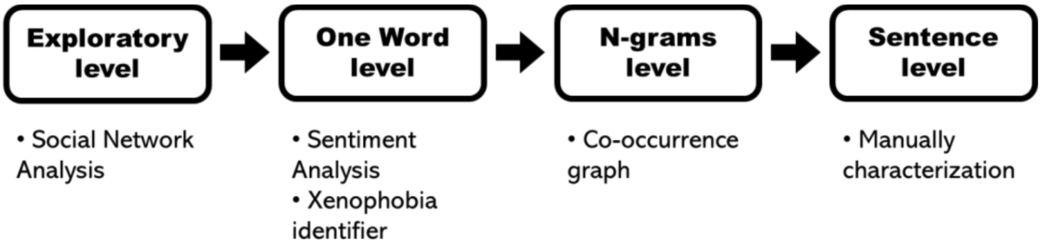
JOSEPH MARTÍNEZ

Note:

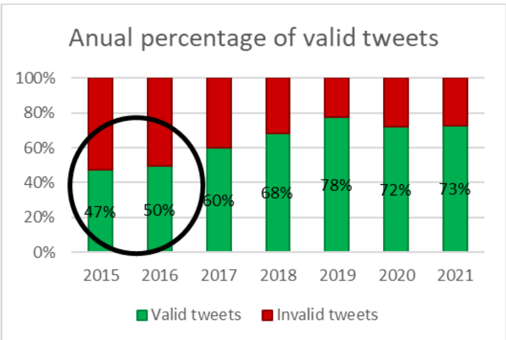
The manual for the tools and codes used in this report is in a different file.

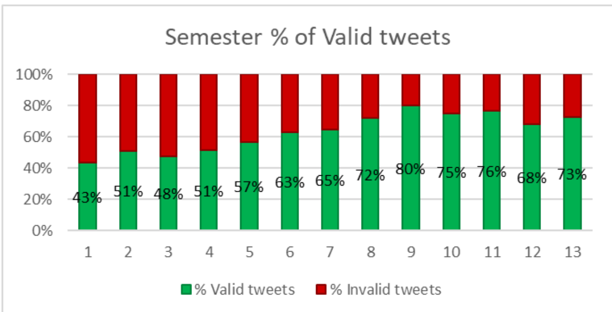
Content

To analyze the conversation the approaches were divided in three branches, that refers to the level of complexity of the used methodologies. These are represented below. First, no texts will be analyzed, only the entities related to the conversation, tweets, users, hashtags, media, and its relationships; Social Network Analysis will be used due the visual simplicity. Second, in a text will be analyzed in a word level, this means that these words will be observed individually, not having into account its combinations. Third, the words combinations or n-grams will be considered to have a wider context. Finally using a manual characterization will be considered the whole tweet. These methodologies are in the next sections.



After the manual validation of the 4683 tweets, we obtained that 2990 of them were about our study subject (63.8%). As shown in the graphs below, 2015 and 2016 were the years with less percentage of validity, this is because in those years there were other trending topics about migrations, but not the one that is studied here. Some of them were illegal smuggling of migrants, not Venezuelan migrants, 'intern migrants' better known as 'desplazados' and migration of Latin-Americans to USA.

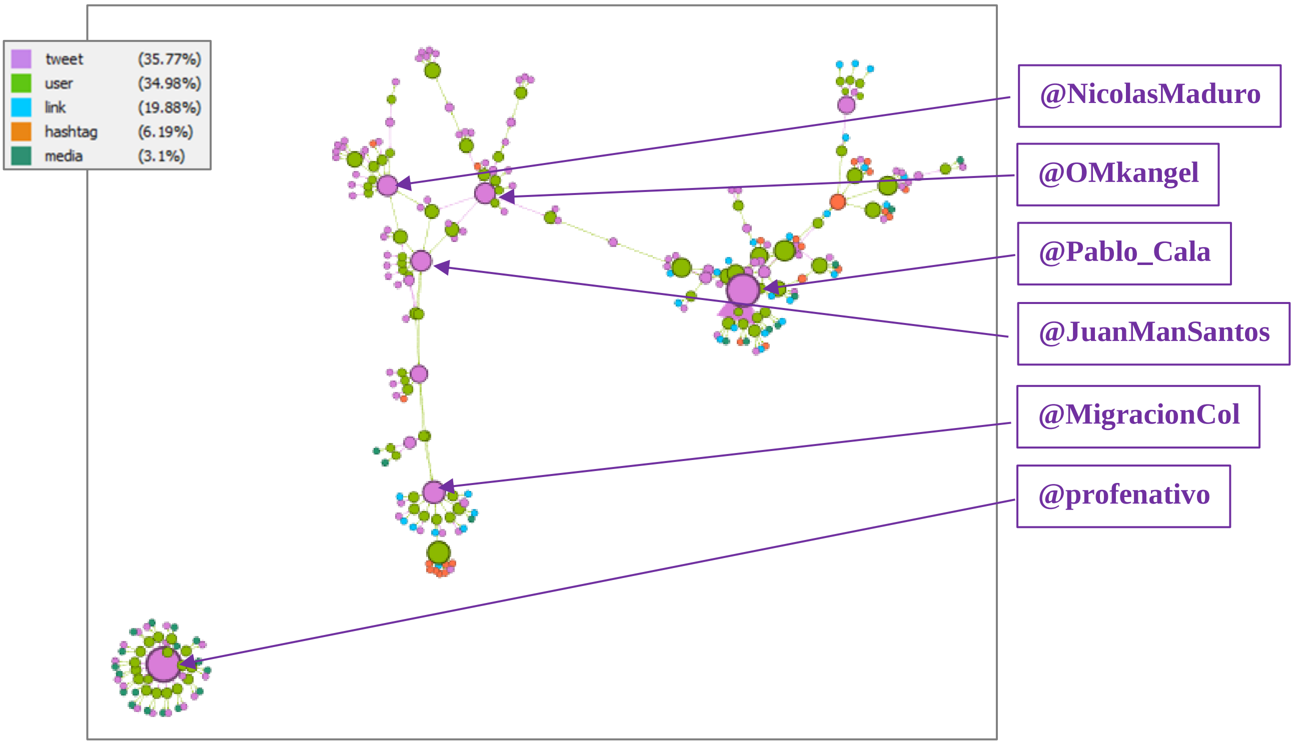




1. Social Network Analysis

In this part will be analyzed the associated Social Network of the conversation per year and at the end a conclusion of the six years of study.

a. 2015

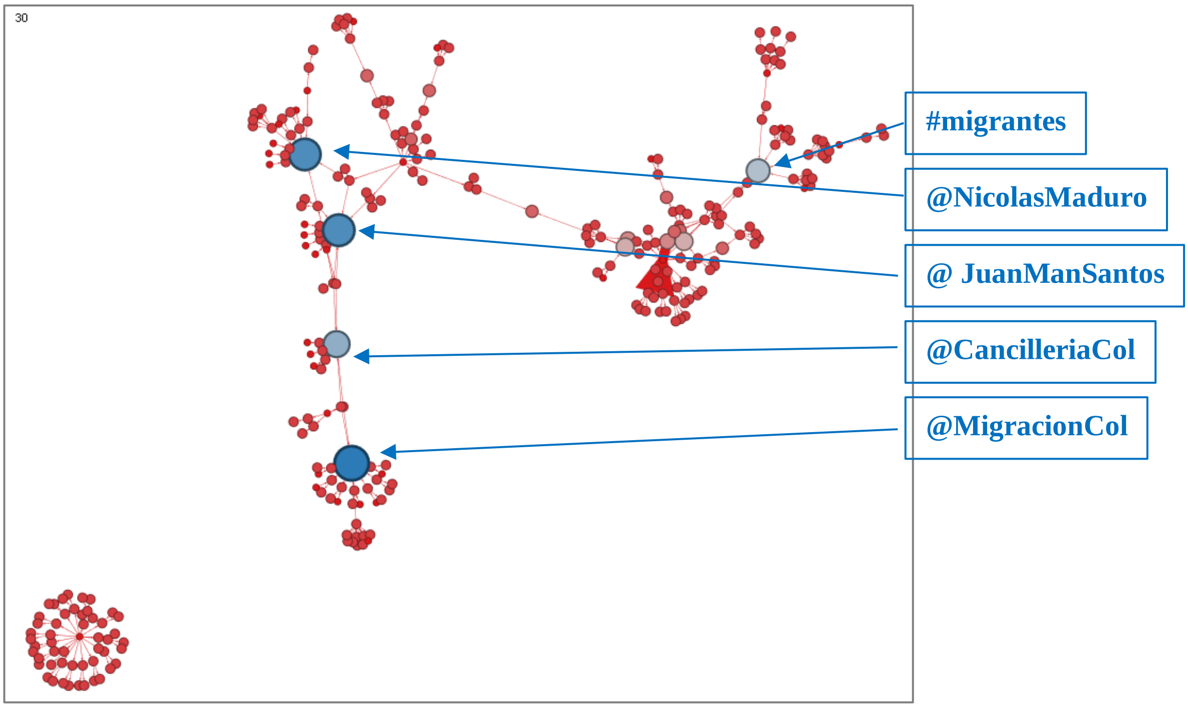


In this visualization can be identified the more relevant entities for the specific year. For example, using the degree we can identify that the users shown above (@Pablo_Cala, @aureliobambiloni, @MigracionCol, @profenativo) are the most important in terms of the connections they have with other entities. This visualization was made according to the entity type but considering for the size the Degree value. This are the ten entities with the higher Degree values:

Entity	Degree value
@profenativo	17
@Pablo_Cala	16
661557443403587584	10
@MigracionCol	10
579067195860234240	9
@JuanManSantos	9
@NicolasMaduro	9
@OMkangel	9
644942452076511232	8
667097048899035136	8

Additional metrics will be used to have other relevant entities with diverse criteria.

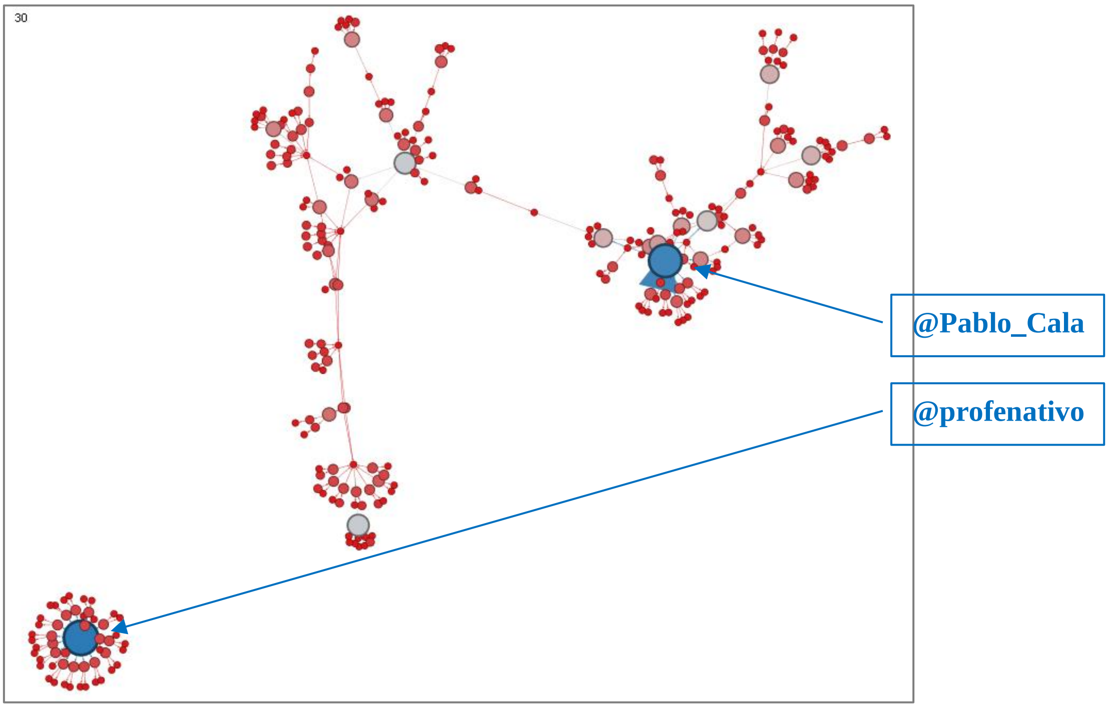
In-Degree



With this metric, we obtain that @MigracionCol has the higher value, this means that in this year most of the people tweet or mention them, gaining importance. There are also other with lower values but also important like #migrantes, @CancilleriaCol, @JuanManSantos and @NicolasMaduro. This are the ten entities with the higher In-Degree values:

Entity	In-Degree value
@MigracionCol	10
@JuanManSantos	9
@NicolasMaduro	9
@CancilleriaCol	7
#migrantes	6
@NoticiasCaracol	5
@elespectador	4
@CDMSaltillo	4
@CIDH	4
@SantaFe	4

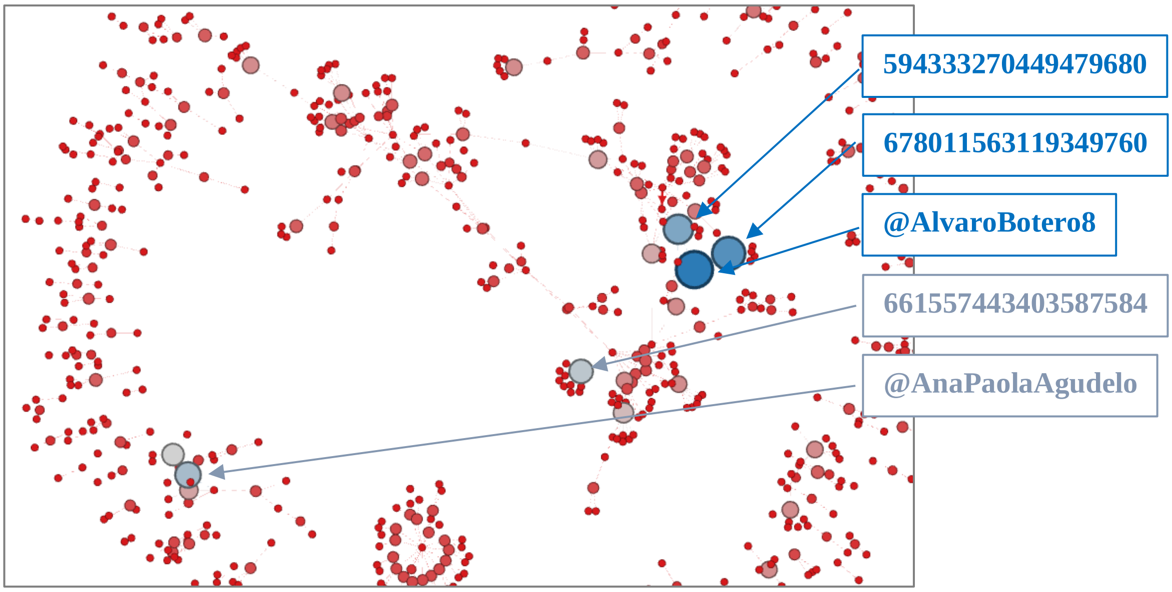
Out-Degree



With this metric, we obtain that @profenativo has the higher value, this means that in this year the account tweeted and retweeted very often, being a content generator in the conversation. There are also other with lower values but also important like @Pablo_Cala, @OMkangek, and @Bettozuleta. This are the ten entities with the higher Out-Degree values:

Entity	Out-Degree value
@profenativo	17
@Pablo_Cala	16
661557443403587584	9
@OMkangel	9
579067195860234240	8
644942452076511232	7
667097048899035136	7
@Bettozuleta	7
594333270449479680	6
594332725240279040	6

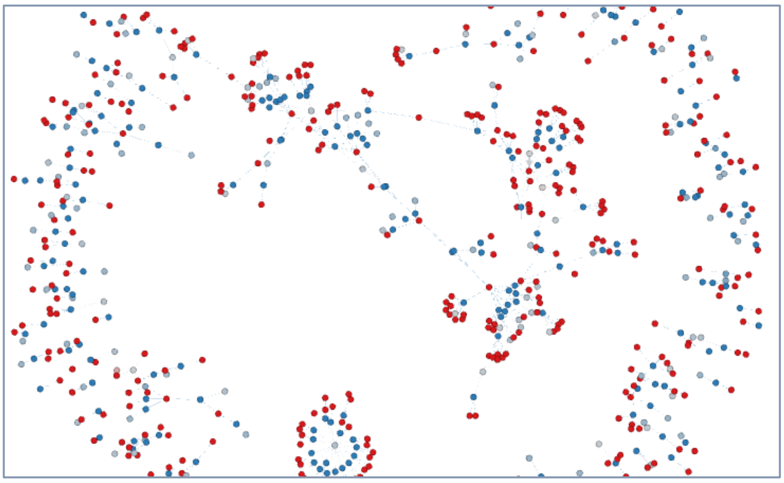
Betweenness centrality



With this metric, we obtain that @AlvaroBotero8 has the higher value, this means that in this year the information passed often through this account, being a content connector in the conversation. There are also other with lower values but also important like 594333270449479680, @678011563119349760, and @AnaPaolaAgudelo. This are the ten entities with the higher Betweenness centrality values:

Entity	Betweenness centrality
@AlvaroBotero8	16
678011563119349760	14
594333270449479680	12
@AnaPaolaAgudelo	10
661557443403587584	9
610279095704887296	8
667097048899035136	7
579067195860234240	6.25
661993551958564864	6
644942452076511232	5.666666667

Closeness centrality



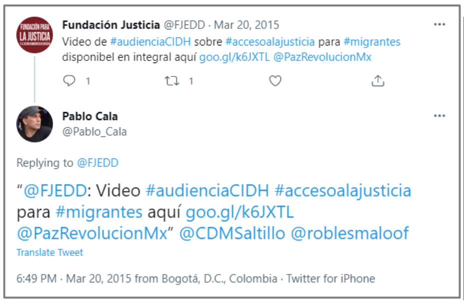
For this measure, the result was very disperse considering a high quantity of nodes with different values, making difficult the identification of important entities.

As a summary, this are the more relevant entities per every metric. Additionally colored for the entities that appears two times and three times.

Degree	In-Degree	Out-degree	Betweenness centrality
@profenativo	@MigracionCol	@profenativo	@AlvaroBotero8
@Pablo_Cala	@JuanManSantos	@Pablo_Cala	678011563119349760
661557443403587584	@NicolasMaduro	661557443403587584	594333270449479680
@MigracionCol	@CancilleriaCol	@OMkangel	@AnaPaolaAgudelo
579067195860234240	#migrantes	579067195860234240	661557443403587584
@JuanManSantos	@NoticiasCaracol	644942452076511232	610279095704887296
@NicolasMaduro	@elespectador	667097048899035136	667097048899035136
@OMkangel	@CDMSaltillo	@Bettozuleta	579067195860234240
644942452076511232	@CIDH	594333270449479680	661993551958564864
667097048899035136	@SantaFe	594332725240279040	644942452076511232

With this, we can identify that both in-degree and out-degree contributes in similar proportions to the degree, as three entities are relevant for the in-degree and degree and five entities are relevant for the out-degree and degree.

It is important to notice that the tweets 661557443403587584, 579067195860234240, 644942452076511232, and 667097048899035136 have high values in 3 of 4 metrics, this shows its great relevance, they create and apport content to the network and with this they act as a bridge to connect the information. These are shown respectively:


Fundación Justicia @FJEDD · Mar 20, 2015
Video de #audienciaCIDH sobre #accesoalajusticia para #migrantes disponibel en integral aquí goo.gl/k6JXTL @PazRevolucionMx

Pablo Cala
@Pablo_Cala

Replying to @FJEDD

"@FJEDD: Video #audienciaCIDH #accesoalajusticia para #migrantes aquí goo.gl/k6JXTL @PazRevolucionMx" @CDMSaltillo @roblesmaloof
Translate Tweet

6:49 PM · Mar 20, 2015 from Bogotá, D.C., Colombia · Twitter for iPhone

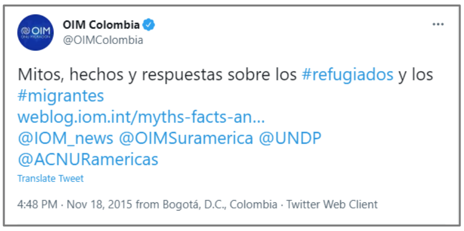

Pablo Cala
@Pablo_Cala

Urge protección internacional a migrantes!!!
@CIDH @hivos @PlataformaDesc @epigmenioibarra @alba_tv @IvanCepedaCast twitter.com/pcslatin/statu...
Translate Tweet

This Tweet is unavailable.

1:33 PM · Sep 18, 2015 from Bogotá, D.C., Colombia · Twitter for iPhone

4 Retweets


OIM Colombia
@OIMColombia

Mitos, hechos y respuestas sobre los #refugiados y los #migrantes
weblog.iom.int/myths-facts-an...
@IOM_news @OIMSuramerica @UNDP @ACNURamericas
Translate Tweet

4:48 PM · Nov 18, 2015 from Bogotá, D.C., Colombia · Twitter Web Client

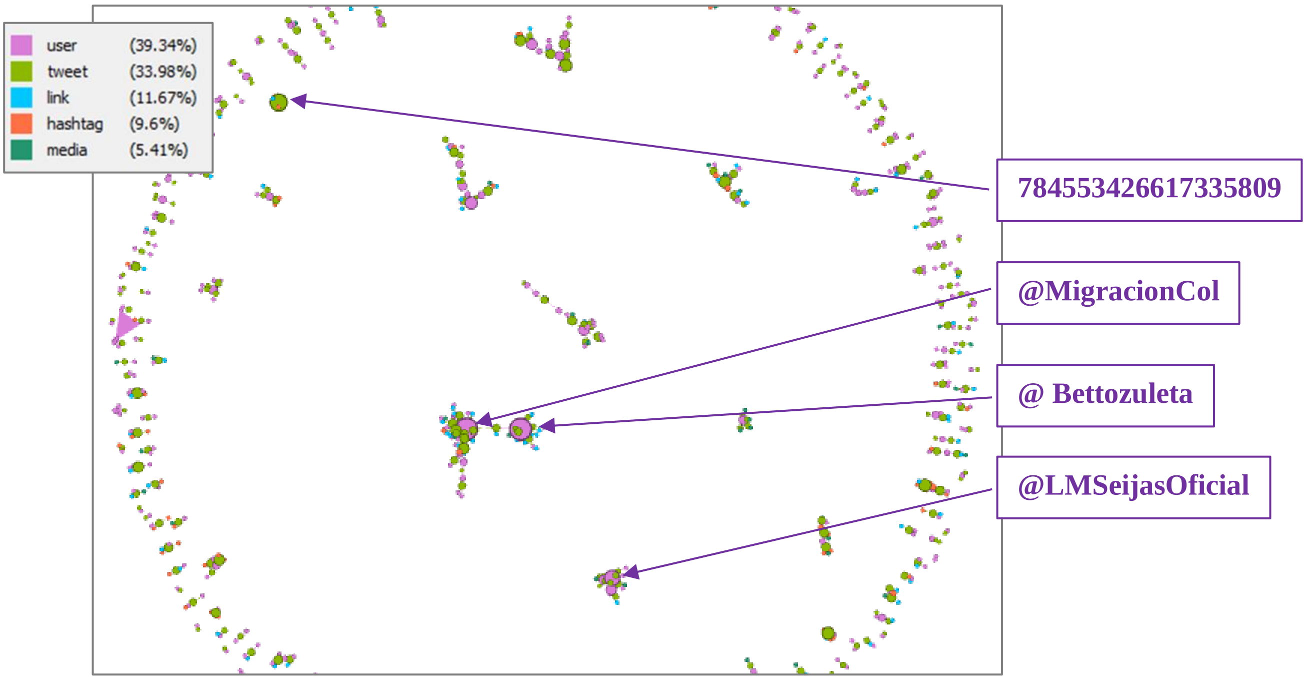
Lastly, using the Gephi's PageRank[1] methodology we have the list of the ten entities with the higher relevance according to the rank metric:

Degree
@NicolasMaduro
@JuanManSantos
@NoticiasCaracol
@MigracionCol
@CancilleriaCol

[1] Algorithm taken from: Sergey Brin, Lawrence Page, The Anatomy of a Large-Scale Hypertextual Web Search Engine, in Proceedings of the seventh International Conference on the World Wide Web (WWW1998):107-117

@felipe_gc10
636740021148626944
@SantaFe
@elespectador
@Hora20

b. 2016

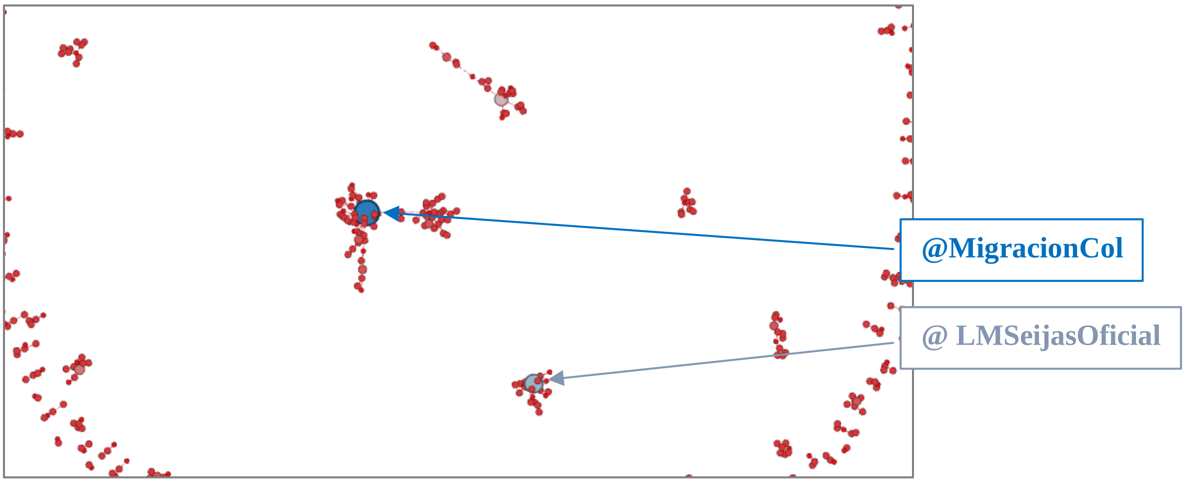


In this visualization can be identified the more relevant entities for the specific year. For example, using the degree we can identify that the users shown above @Bettozuleta, @MigracionCol, and @LMSeijasOficial) are the most important in terms of the connections they have with other entities. This visualization was made according to the entity type but considering for the size the Degree value. This are the ten entities with the higher Degree values:

Entity	Degree value
@MigracionCol	12
@Bettozuleta	12
784553426617335809	9
@LMSeijasOficial	8
@caladofer17	6
756054401811349504	6
780393919678906368	6
791801368654778368	6
805340426089668608	6
@JuanManSantos	5

Additional metrics will be used to have other relevant entities with diverse criteria.

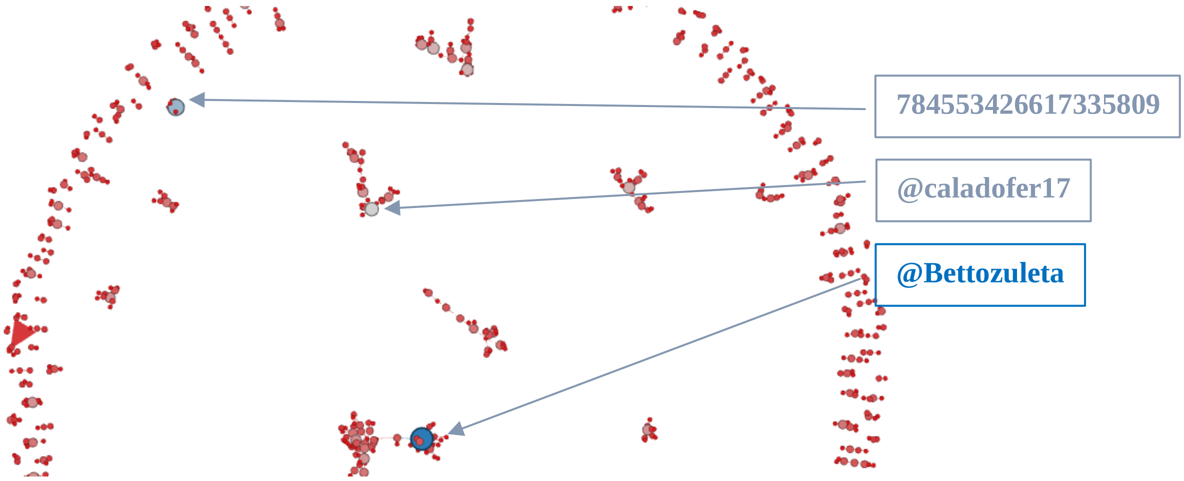
In-Degree



With this metric, we obtain that @MigracionCol has the higher value, this means that in this year most of the people tweet or mention them, gaining importance. There are also other with lower values but also important like @LMSeijasOficial. This are the ten entities with the higher In-Degree values:

Entity	In-Degree value
@MigracionCol	12
@LMSeijasOficial	8
@JuanManSantos	5
@NicolasMaduro	4
@RevistaSemana	4
@SantaFe	3
@NoticiasRCN	3
@tripleCIbarguen	3
@OIMColombia	2
@CancilleriaCol	2

Out-Degree

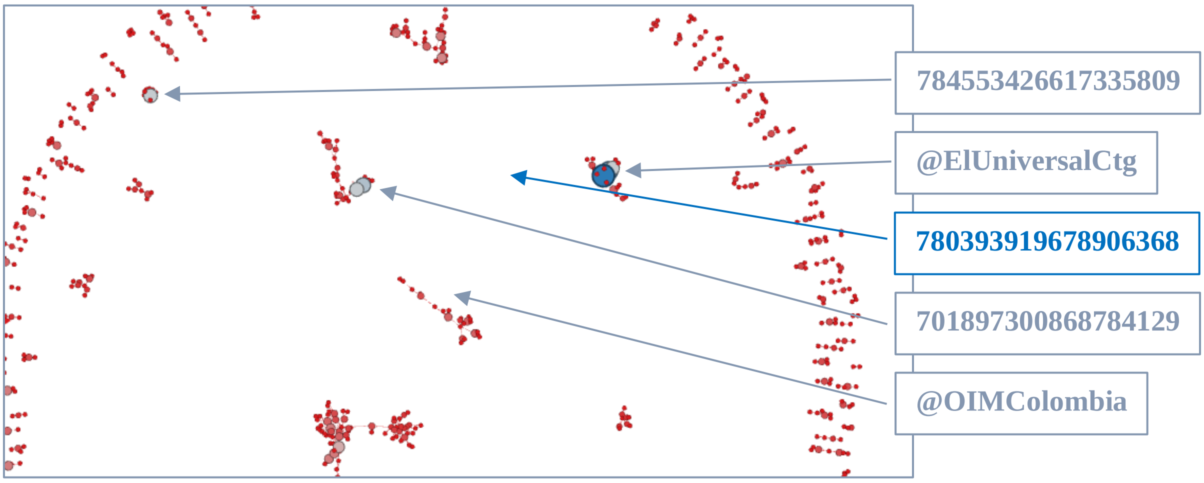


With this metric, we obtain that @Bettozuleta has the higher value, this means that in this year the account tweeted and retweeted very often, being a content generator in the conversation. There are also other with lower values but also important like 797277036691787776 and @caladofer17. This are the ten entities with the higher Out-Degree values:

Entity	Out-Degree value

@Bettozuleta	12
784553426617335809	8
@caladofer17	6
756054401811349504	5
780393919678906368	5
791801368654778368	5
805340426089668608	5
@MateoRod23	5
@erodrigueznql	5
710147707021729792	4

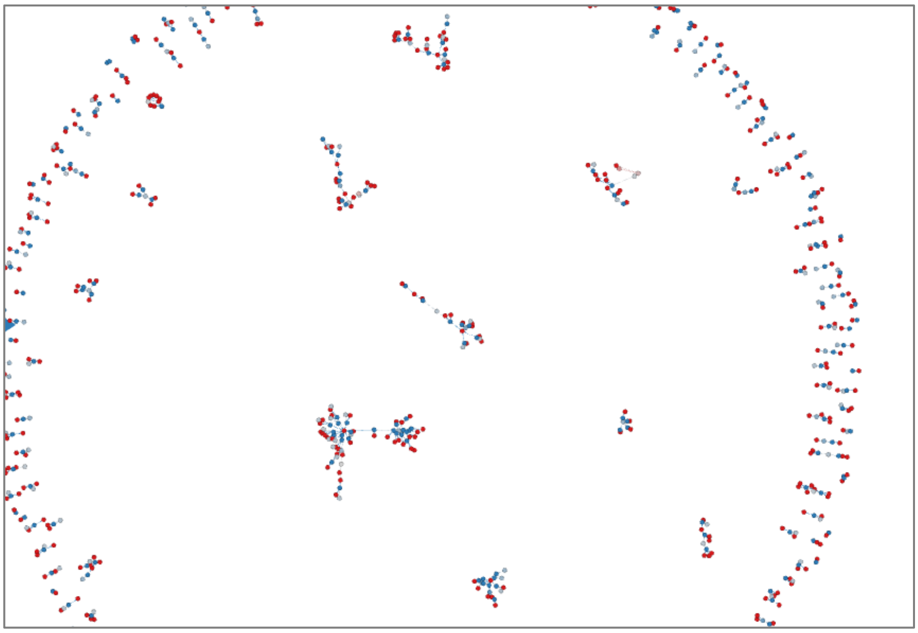
Betweenness centrality



With this metric, we obtain that 780393919678906368 has the higher value, this means that in this year the information passed often through this account, being a content connector in the conversation. There are also other with lower values but also important like @ElUniversalCtg, 701897300868784129, and @OIMColombia. This are the ten entities with the higher Betweenness centrality values:

Entity	Betweenness centrality
780393919678906368	15
@ElUniversalCtg	12
701897300868784129	9
@OIMColombia	8
777978609063591936	8
784553426617335809	8
746413280277708800	6
756054401811349504	5
791801368654778368	4.5
805340426089668608	4.5

Closeness centrality

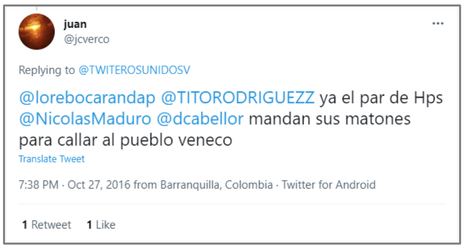


For this measure, the result was very disperse considering a high quantity of nodes with different values, making difficult the identification of important entities.

As a summary, this are the more relevant entities per every metric. Additionally colored for the entities that appears two times and three times.

Degree	In-Degree	Out-degree	Betweenness centrality
@MigracionCol	@MigracionCol	@Bettozuleta	780393919678906368
@Bettozuleta	@LMSeijasOficial	784553426617335809	@ElUniversalCtg
784553426617335809	@JuanManSantos	@caladofer17	701897300868784129
@LMSeijasOficial	@NicolasMaduro	756054401811349504	@OIMColombia
@caladofer17	@RevistaSemana	780393919678906368	777978609063591936
756054401811349504	@SantaFe	791801368654778368	784553426617335809
780393919678906368	@NoticiasRCN	805340426089668608	746413280277708800
791801368654778368	@tripleCIbarguen	@MateoRod23	756054401811349504
805340426089668608	@OIMColombia	@erodriguezngl	791801368654778368
@JuanManSantos	@CancilleriaCol	710147707021729792	805340426089668608

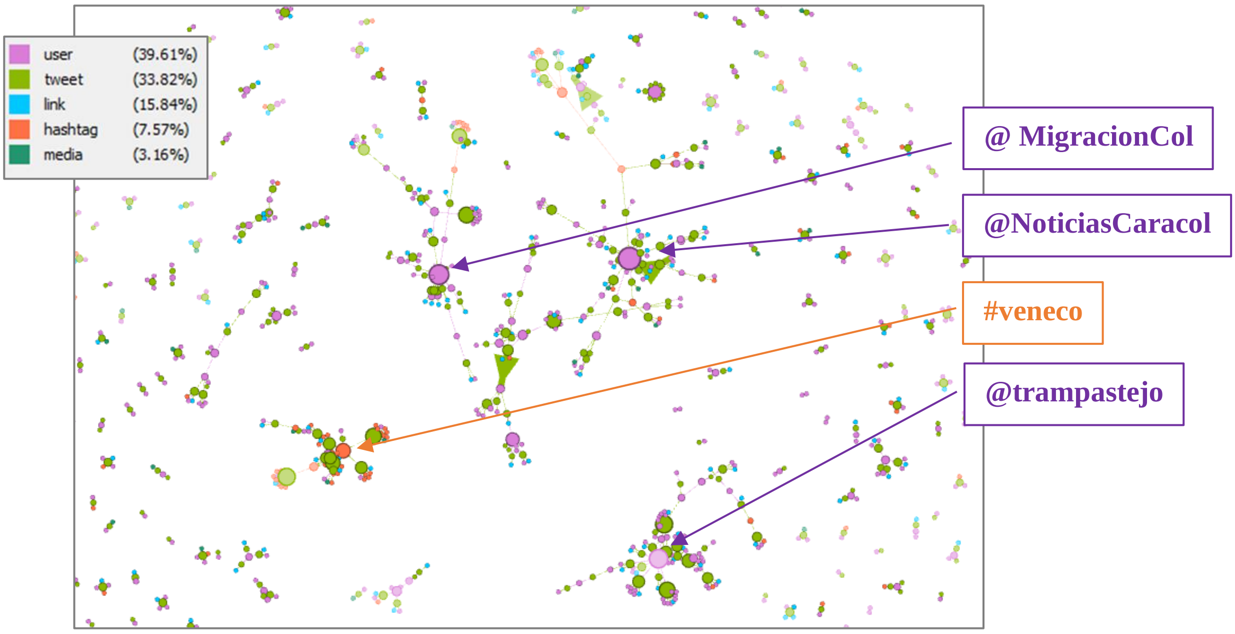
With this, we can identify that both in-degree and out-degree contributes in similar proportions to the degree, as three entities are relevant for the in-degree and degree and five entities are relevant for the out-degree and degree. It is important to notice that the tweets 784553426617335809, 756054401811349504, 780393919678906368, 791801368654778368, and 805340426089668608 have high values in 3 of 4 metrics, this shows its great relevance, they create and apport content to the network and with this they act as a bridge to connect the information. These are shown:



Lastly, using the Gephi's PageRank methodology we have the list of the ten entities with the higher relevance according to the rank metric:

Degree
@LMSeijasOficial
@MigracionCol
@NANCARDENAS_
733448545165021184
@manducogh
733091791805779969
@RevistaSemana
@PadreJosePalmar
@NoticiasRCN
@JuanManSantos

c. 2017

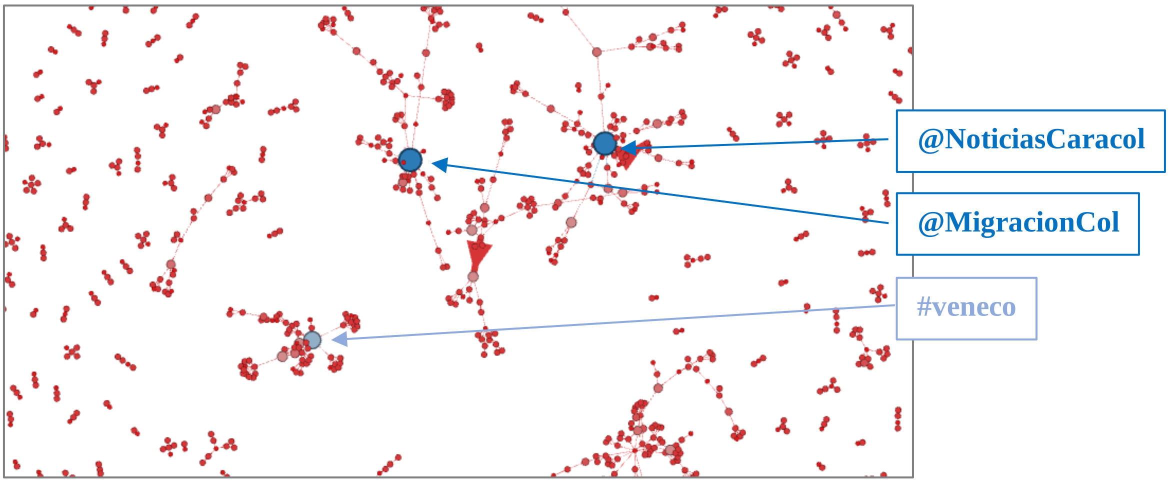


In this visualization can be identified the more relevant entities for the specific year. For example, using the degree we can identify that the users shown above @NoticiasCaracol, @trampastejo, and @MigracionCol), and the hashtag #veneco are the most important in terms of the connections they have with other entities. This visualization was made according to the entity type but considering for the size the Degree value. This are the ten entities with the higher Degree values:

Entity	Degree value
@MigracionCol	15
@NoticiasCaracol	13
@trampastejo	12
865263940556738561	11
873927438815907840	11
865385529822642178	10
886976706422202369	10
919269530991497216	10
#veneco	9
845290876003909632	9

Additional metrics will be used to have other relevant entities with diverse criteria.

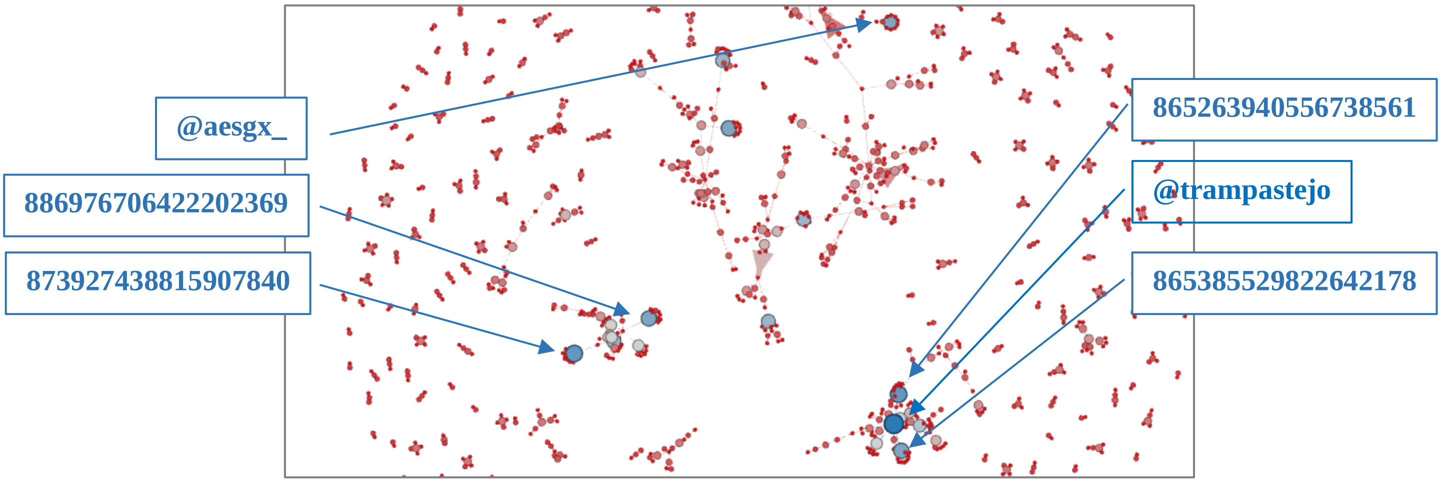
In-Degree



With this metric, we obtain that @NoticiasCaracol has the higher value, this means that in this year most of the people tweet or mention them, gaining importance. There are also other with lower values but also important like @MigracionCol and #veneco. This are the ten entities with the higher In-Degree values:

Entity	In-Degree value
@MigracionCol	13
@NoticiasCaracol	13
#veneco	9
@elespectador	5
#USA	5
#Colombia	5
@JuanManSantos	4
@NicolasMaduro	4
#colombia	4
@AlvaroUribeVel	4

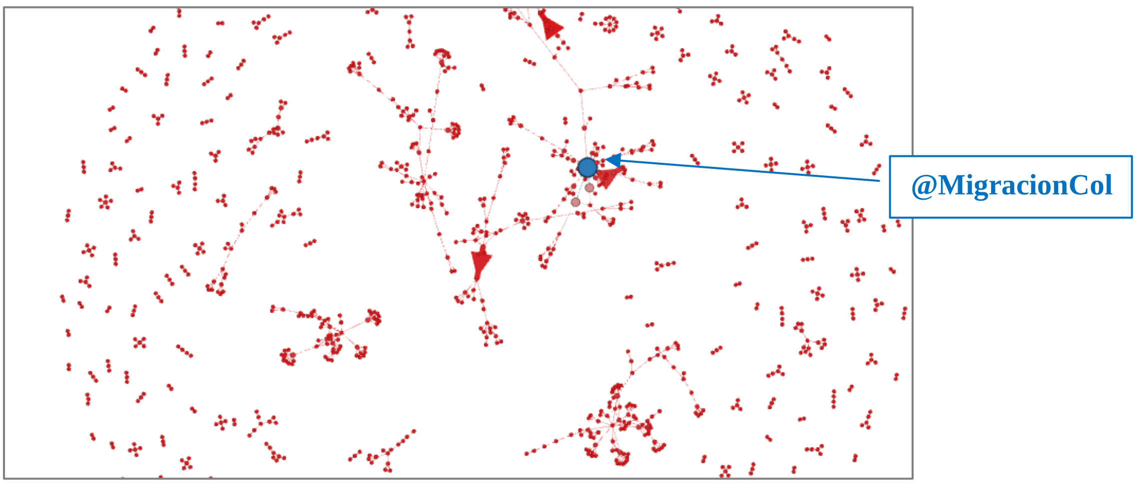
Out-Degree



With this metric, we obtain that @trampastejo has the higher value, this means that in this year the account tweeted and retweeted very often, being a content generator in the conversation. There are also other with lower values but also important like @aesgx_. In this particular case we obtained that many nodes has high out-degree values, this shows that more people is tweeting and more content is generated.   This are the ten entities with the higher Out-Degree values:

Entity	Out-Degree value
@trampastejo	12
865263940556738561	10
873927438815907840	10
865385529822642178	9
886976706422202369	9
919269530991497216	9
@aesgx_	9
845290876003909632	8
842212452955787264	8
887628019417960448	8

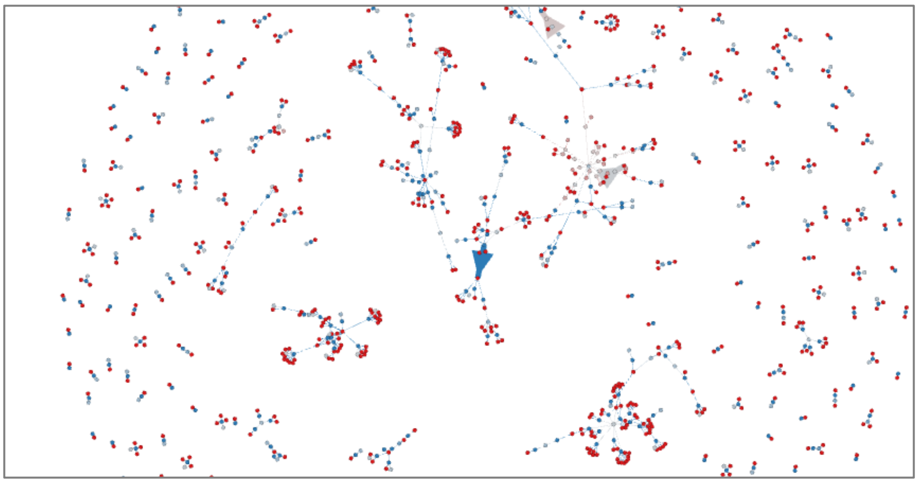
Betweenness centrality



With this metric, we obtain that @MigracionCol has the higher value, this means that in this year the information passed often through this account, being a content connector in the conversation. This are the ten entities with the higher Betweenness centrality values:

Entity	Betweenness centrality
@MigracionCol	125
872076747172433920	39
871929036737576962	39
873927438815907840	10
836307884543774720	9
865385529822642178	9
886976706422202369	9
898734324010295296	9
919269530991497216	9
865263940556738561	8.83

Closeness centrality



For this measure, the result was very disperse considering a high quantity of nodes with different values, making difficult the identification of important entities.

As a summary, this are the more relevant entities per every metric. Additionally colored for the entities that appears two times and three times.

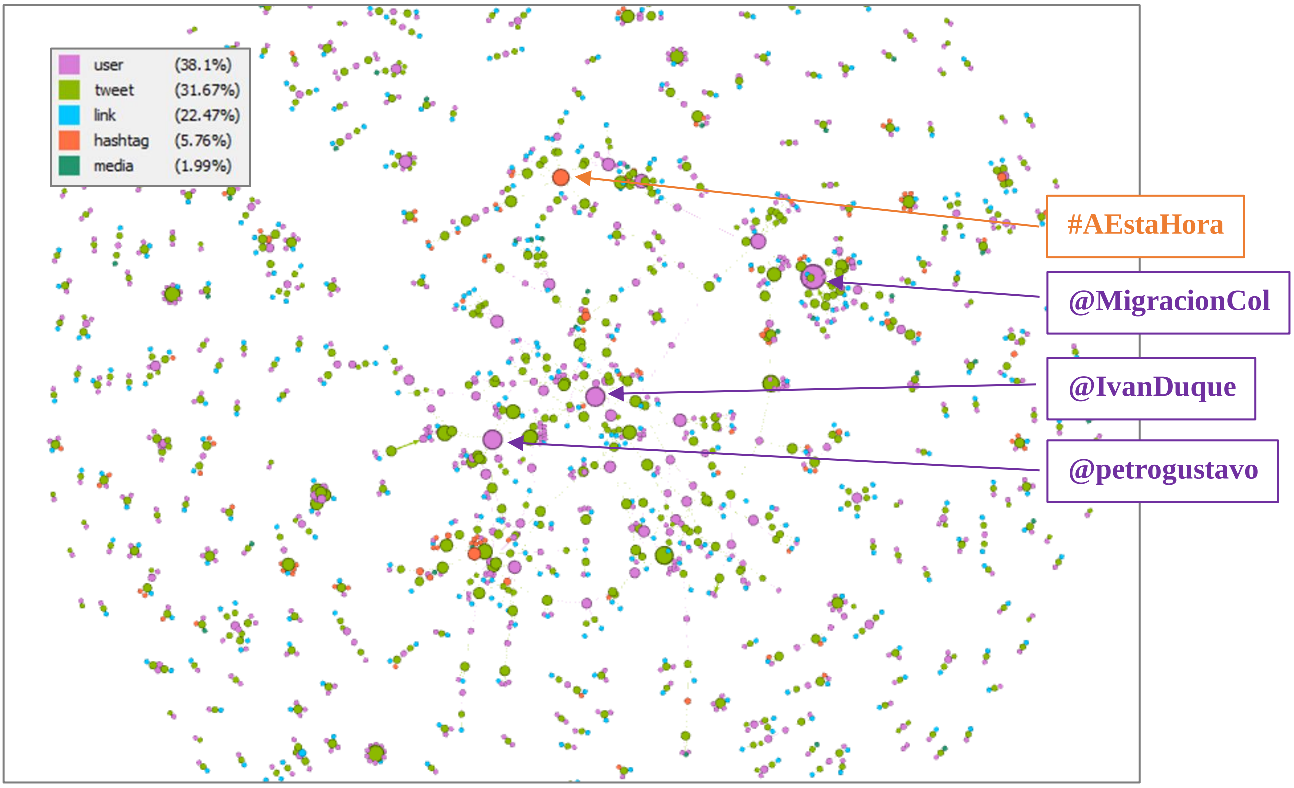
Degree	In-Degree	Out-degree	Betweenness centrality
@MigracionCol	@MigracionCol	@trampastejo	@MigracionCol
@NoticiasCaracol	@NoticiasCaracol	865263940556738561	872076747172433920
@trampastejo	#veneco	873927438815907840	871929036737576962
865263940556738561	@elespectador	865385529822642178	873927438815907840
873927438815907840	#USA	886976706422202369	836307884543774720
865385529822642178	#Colombia	919269530991497216	865385529822642178
886976706422202369	@JuanManSantos	@aesgx_	886976706422202369
919269530991497216	@NicolasMaduro	845290876003909632	898734324010295296
#veneco	#colombia	842212452955787264	919269530991497216
845290876003909632	@AlvaroUribeVel	887628019417960448	865263940556738561

With this, we can identify that the out-degree value contributes more in the degree than the in-degree. Making the entities with higher out-degree values also the ones with higher degree values, this means, the entities that creates or contains content. However, there are exceptions like @NoticiasCaracol and #veneco, with high in-degree values that reflect how many entities mentions him. Here, for *#veneco* is important to keep in mind that this pejorative term started to gain popularity and use. It is important to notice that @MigracionCol has high values in 3 of 4 metrics, this shows its great relevance, they are mentioned in plenty of tweets and they act as a bridge to connect the information. In this year, we can see that the conversation is growing about the migration, considering the importance of @MigracionCol and *#veneco* in this year. These are shown:

Lastly, using the Gephi's PageRank methodology we have the list of the ten entities with the higher relevance according to the rank metric:

Degree
@NoticiasCaracol
@MigracionCol
@BOG_ELDORADO
#SomosMigración
872076747172433920
871929036737576962
#USA
@German_Vargas
#veneco
@DrodriguezVen

d. 2018



In this visualization can be identified the more relevant entities for the specific year. For example, using the degree we can identify that the users shown above @IvanDuque, @petrogustavo, and @MigracionCol) and the hashtag #AEstaHora are the most important in terms of the connections they have with other entities. #AEstaHora reflects that for this year many events were tweeted around the Venezuelan migration.

This visualization was made according to the entity type but considering for the size the Degree value.

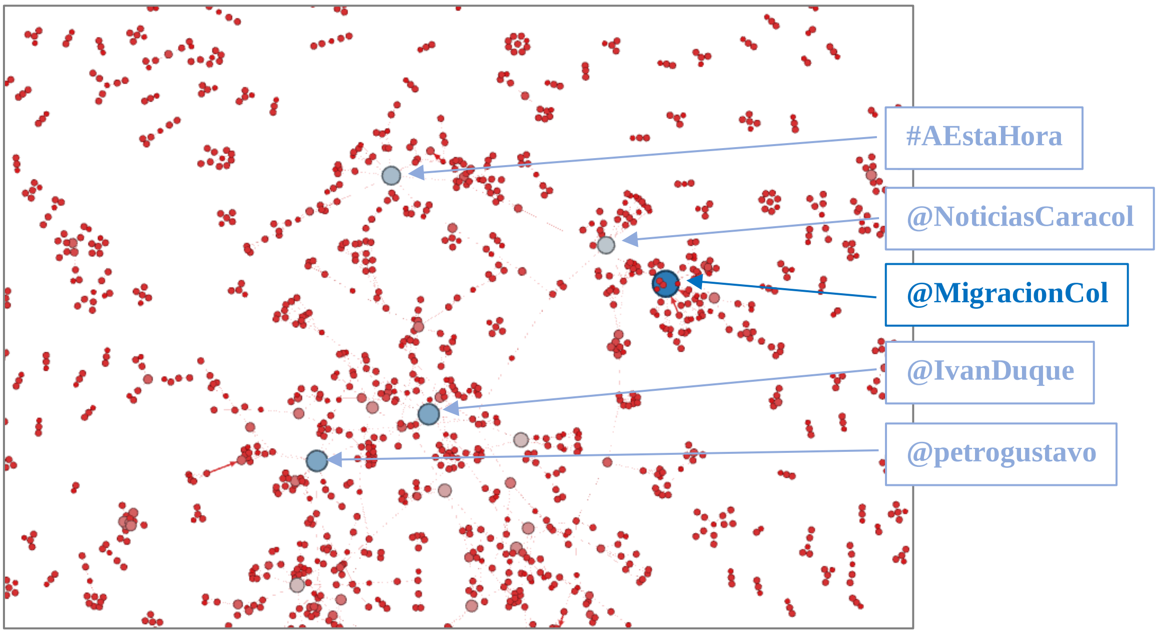
Considering that the significant number of nodes with high degree, this are the twelve entities with the higher Degree values:

Entity	Degree value
@MigracionCol	16
@petrogustavo	12
@IvanDuque	12
1056027033036341249	11
#AEstaHora	10
975382872575168513	10
1067787816443281409	10
@NoticiasCaracol	9
956020868236431360	9
996853142896697344	9
1067480113766375425	9
1077967989192953859	9
968236810026577921	8
@AdrianaMatizTol	8
1044561110316306433	8
1067845278147981312	8
1067240117847425024	8

@AlvaroUribeVel	7
#Colombia	7
@achuri_pablo	7

Additional metrics will be used to have other relevant entities with diverse criteria.

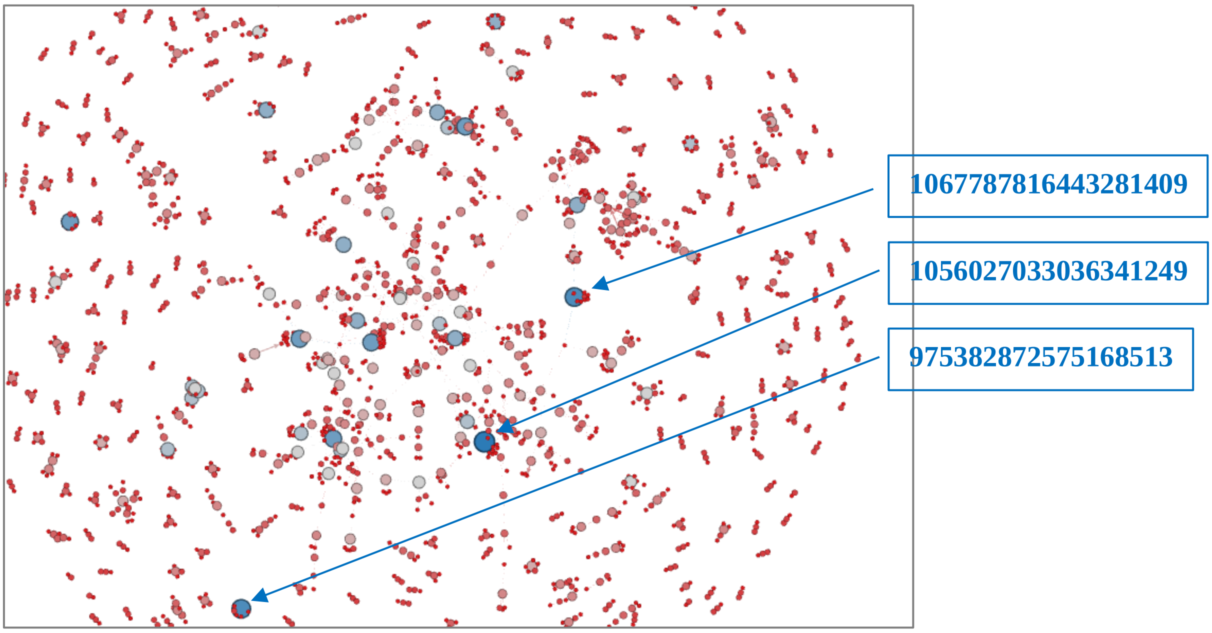
In-Degree



With this metric, we obtain that @MigracionCol has the higher value, this means that in this year most of the people tweet or mention them, gaining importance. There are also other with lower values but also important like @NoticiasCaracol, @IvanDuque, @petrogustavo and #AEstaHora. This are the twelve entities with the higher In-Degree values:

Entity	In-Degree value
@MigracionCol	16
@petrogustavo	12
@IvanDuque	12
#AEstaHora	10
@NoticiasCaracol	9
@AlvaroUribeVel	7
#Colombia	7
@NicolasMaduro	6
@JuanManSantos	5
@ACNURamericas	5
@CarlosHolmesTru	5
@CancilleriaVE	5
@mluciaramirez	4
@ELTIEMPO	4
@elespectador	4
@BOG_ELDORADO	4
@BluRadioCo	4
@MillosFCoficial	4
#Venezuela	4
@AndresPastrana_	4

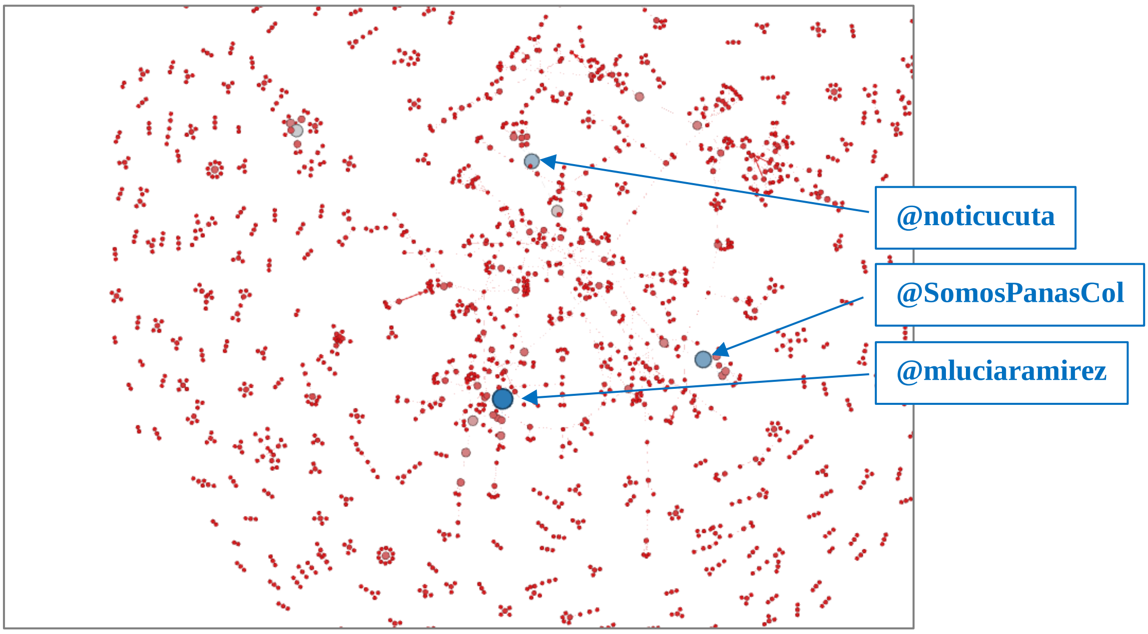
Out-Degree



With this metric, we obtain that 1056027033036341249, @UNODCColombia, and @ClementeCR2016 have the higher value, this means that in this year the account tweeted and retweeted very often, being a content generator in the conversation. There are also other ones with lower values, but still considerably important. This are the twelve entities with the higher Out-Degree values:

Entity	Out-Degree value
1056027033036341249	10
975382872575168513	9
1067787816443281409	9
956020868236431360	8
996853142896697344	8
1067480113766375425	8
1077967989192953859	8
@AdrianaMatizTol	8
968236810026577921	7
1044561110316306433	7
1067845278147981312	7
1067240117847425024	7
@achuri_pablo	7
@ExtraNoticiasCo	7
@UNODCColombia	7
956181322996834304	6
954889077597392896	6
953442756197691392	6
1020734948662226944	6
1034540722240847873	6

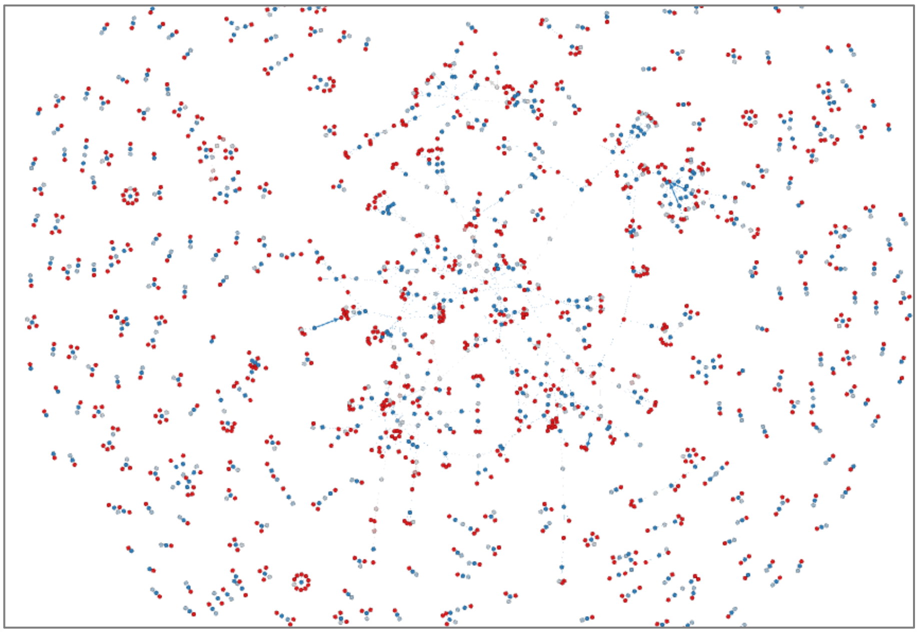
Betweenness centrality



With this metric, we obtain that @mluciaramirez has the higher value, this means that in this year the information passed often through this account, being a content connector in the conversation. This are the twelve entities with the higher Betweenness centrality values:

Entity	Betweenness centrality
@mluciaramirez	42
@SomosPanasCol	32
@noticucuta	28
@RadNalCo	22
988773573039939589	19
1075903658854879235	15
@GobiernoBTA	12
987735210358968320	12
1022941957591642112	12
@WFP_es	12
1034239966816161792	11.5
957291961018404864	10
955814298982518785	10
1000034594635354114	10
1021806463608676353	10
1055289115808358402	10
1056027033036341249	9.3
975382872575168513	9
1022534767374413826	9
1067787816443281409	9

Closeness centrality

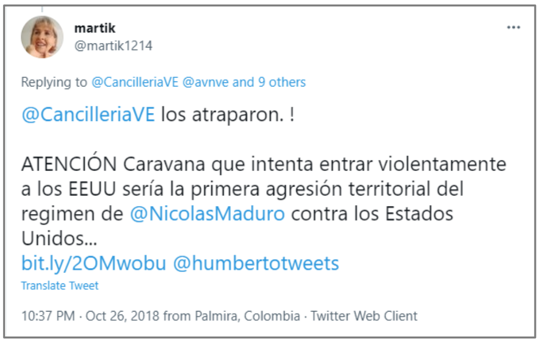


For this measure, the result was very disperse considering a high quantity of nodes with different values, making difficult the identification of important entities.

As a summary, this are the more relevant entities per every metric. Additionally, colored for the entities that appears two times and three times.

Degree	In-Degree	Out-degree	Betweenness centrality
@MigracionCol	@MigracionCol	1056027033036341249	@mluciaramirez
@petrogustavo	@petrogustavo	975382872575168513	@SomosPanasCol
@IvanDuque	@IvanDuque	1067787816443281409	@noticucuta
1056027033036341249	#AEstaHora	956020868236431360	@RadNalCo
#AEstaHora	@NoticiasCaracol	996853142896697344	988773573039939589
975382872575168513	@AlvaroUribeVel	1067480113766375425	1075903658854879235
1067787816443281409	#Colombia	1077967989192953859	@GobiernoBTA
@NoticiasCaracol	@NicolasMaduro	@AdrianaMatizTol	987735210358968320
956020868236431360	@JuanManSantos	968236810026577921	1022941957591642112
996853142896697344	@ACNURamericas	1044561110316306433	@WFP_es
1067480113766375425	@CarlosHolmesTru	1067845278147981312	1034239966816161792
1077967989192953859	@CancilleriaVE	1067240117847425024	957291961018404864
968236810026577921	@mluciaramirez	@achuri_pablo	955814298982518785
@AdrianaMatizTol	@ELTIEMPO	@ExtraNoticiasCo	1000034594635354114
1044561110316306433	@elespectador	@UNODCColombia	1021806463608676353
1067845278147981312	@BOG_ELDORADO	956181322996834304	1055289115808358402
1067240117847425024	@BluRadioCo	954889077597392896	1056027033036341249
@AlvaroUribeVel	@MillosFCoficial	953442756197691392	975382872575168513
#Colombia	#Venezuela	1020734948662226944	1022534767374413826
@achuri_pablo	@AndresPastrana_	1034540722240847873	1067787816443281409

With this, we can identify that both in-degree and out-degree contributes in similar proportions to the degree. As seven entities are relevant for the in-degree and degree and ten entities are relevant for the out-degree and degree. It is important to notice that the tweets 1056027033036341249, 975382872575168513 and 1067787816443281409 have high values in 3 of 4 metrics, this shows its great relevance, they create and apport content to the network and with this they act as a bridge to connect the information.


martik
@martik1214
Replying to @CancilleriaVE @avnve and 9 others
@CancilleriaVE los atraparon. !

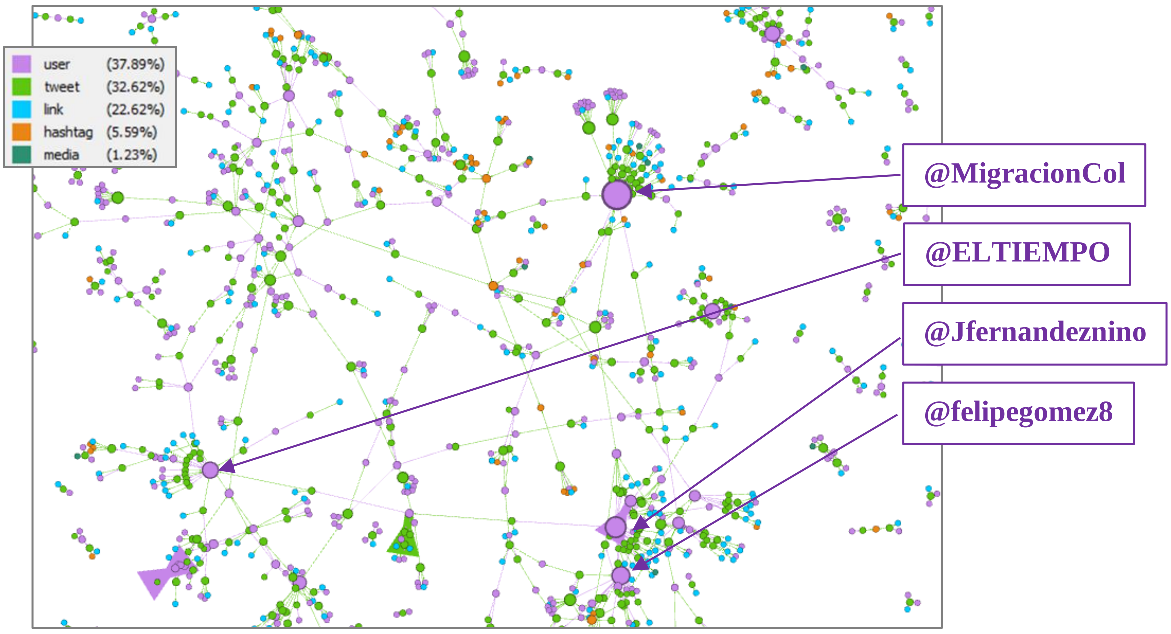
ATENCIÓN Caravana que intenta entrar violentamente a los EEUU sería la primera agresión territorial del regimen de @NicolasMaduro contra los Estados Unidos...
bit.ly/2OMwobu @humbertotweets
Translate Tweet
10:37 PM · Oct 26, 2018 from Palmira, Colombia · Twitter Web Client


@alejo039
@Sativasur039
Replying to @msilvaazul @casaleantonio and 6 others
Estamos con los mas altos índices de inseguridad de todos los tiempos y en gran parte se debe a los migrantes Venezolanos
Translate Tweet
9:46 AM · Mar 18, 2018 from Bogotá, D.C., Colombia · Twitter for Android

Lastly, using the Gephi's PageRank methodology we have the list of the ten entities with the higher relevance according to the rank metric:

Degree
@MigracionCol
@NoticiasCaracol
@AlvaroUribeVel
@IvanDuque
#AEstaHora
@petrogustavo
#Colombia
@JuanManSantos
@CarlosHolmesTru
@MillosFCoficial

e. 2019

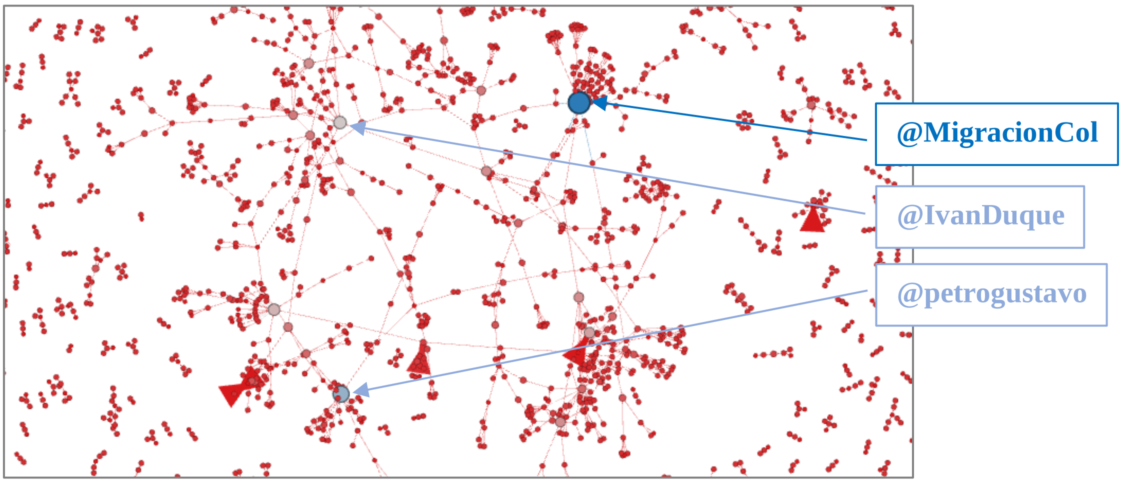


In this visualization can be identified the more relevant entities for the specific year. For example, using the degree we can identify that the users shown above (@ELTIEMPO, @JFernandeznino, @felipemgomez8 and @MigracionCol) are the most important in terms of the connections they have with other entities. This visualization was made according to the entity type but considering for the size the Degree value. This are the twelve entities with the higher Degree values:

Entity	Degree value
@MigracionCol	37
@JFernandeznino	24
@felipemgomez8	20
@ELTIEMPO	17
@aesgx_	16
@JuanNavarrete_	16
@petrogustavo	13
1154228871316856834	11
@UninorteCO	10
@OPSOMS_Col	10
1132401199204970496	10
1132079823982088192	10
1187043585885134851	10
1210633554666897409	10
1208950019627372545	10
@MinSaludCol	9
@IvanDuque	9
@carlosevilma	9
1088875762709643264	9
1131991450638573568	9

Additional metrics will be used to have other relevant entities with diverse criteria.

In-Degree

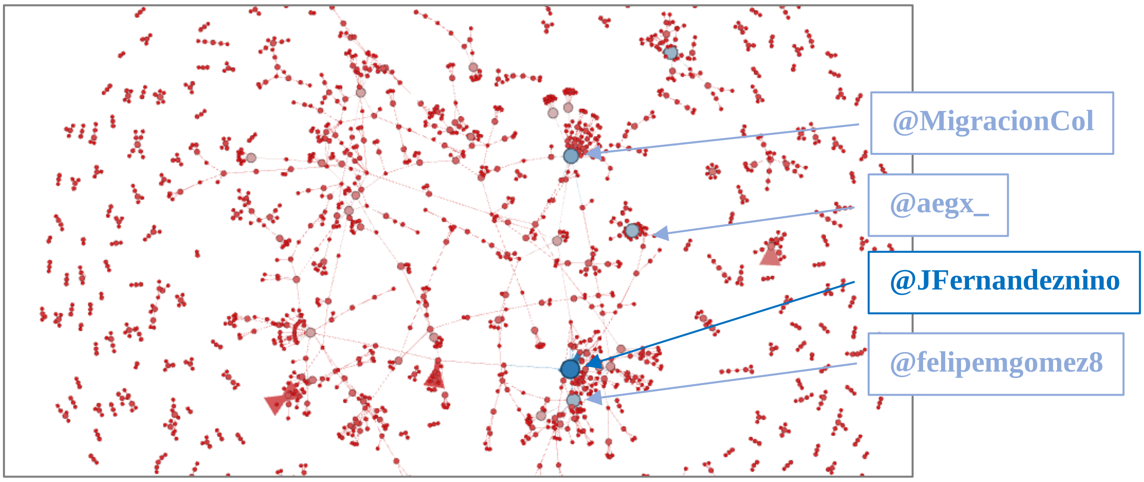


With this metric, we obtain that @MigracionCol has the higher value, this means that in this year most of the people tweet or mention them, gaining importance. There are also other with lower values but also important like @petrogustavo and @IvanDuque. This are the twelve entities with the higher In-Degree values:

Entity	In-Degree value
@MigracionCol	19
@petrogustavo	13
@IvanDuque	9
@ELTIEMPO	8
@UninorteCO	7
@OIMColombia	6
#Colombia	6

@lcvelez	6
#AEstaHora	6
#migrantes	5
@BluRadioCo	5
@lafm	5
@jguaido	5
@felipemgomez8	4
@elespectador	4
@NoticiasCaracol	4
@migravenezuela	4
#Repost	4
@USAID_Colombia	4
@NicolasMaduro	3

Out-Degree

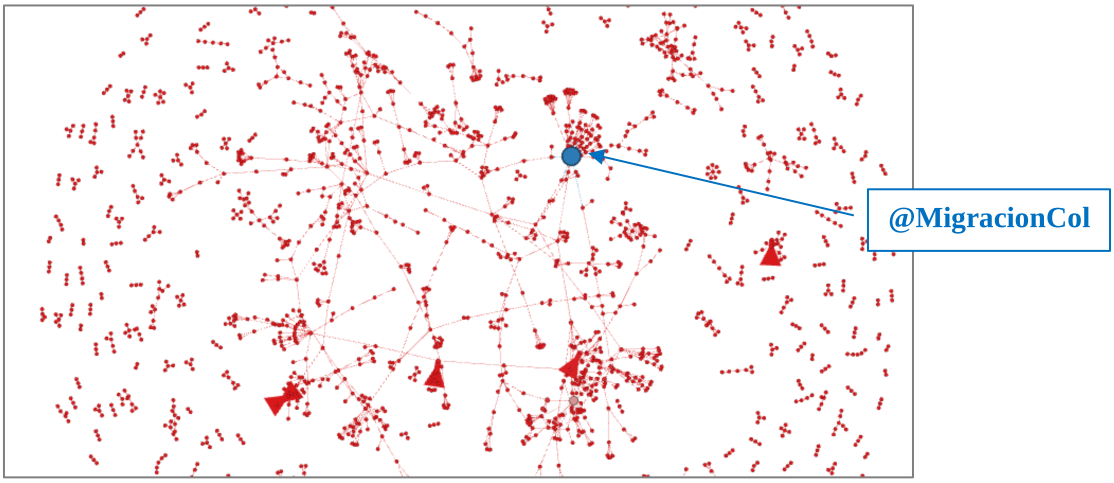


With this metric, we obtain that @JFernandeznino has the higher value, this means that in this year the account tweeted and retweeted very often, being a content generator in the conversation. There are also other with lower values but also important like @ELTIEMPO, @carlosevilma, @felipemgomez8, and @MigracionCol. This are the twelve entities with the higher Out-Degree values:

Entity	Out-Degree value
@JFernandeznino	24
@MigracionCol	18
@felipemgomez8	16
@aesgx_	16
@JuanNavarrete_	16
1154228871316856834	10
@ELTIEMPO	9
1132401199204970496	9
1132079823982088192	9
1187043585885134851	9
1210633554666897409	9
1208950019627372545	9
@carlosevilma	9
@OPSOMS_Col	8
1088875762709643264	8
1131991450638573568	8
1155265790842720256	8

@MinSaludCol	7
1100763592981401600	7
1131955953241350144	7

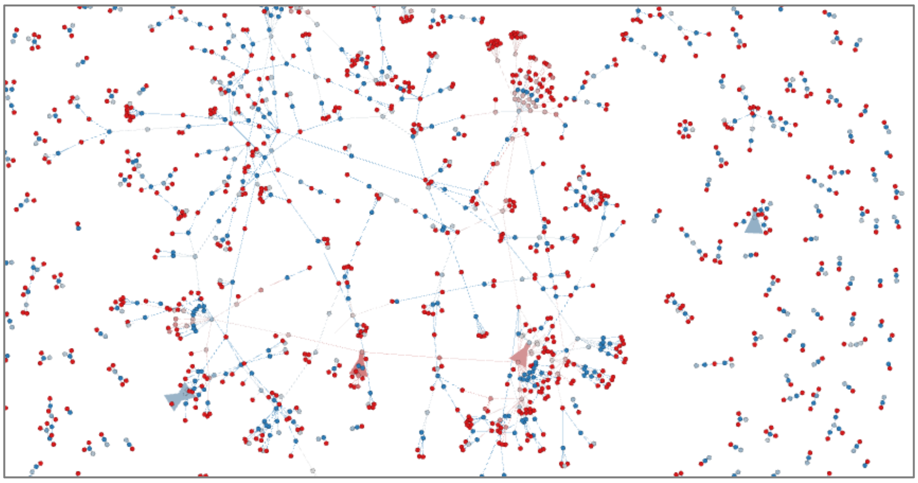
Betweenness centrality



With this metric, we obtain that @MigracionCol has the higher value, this means that in this year the information passed often through this account, being a content connector in the conversation. This are the twelve entities with the higher Betweenness centrality values:

Entity	Betweenness centrality
@MigracionCol	1159
@felipemgomez8	382.5
@ELTIEMPO	324
@OPSOMS_Col	166.5
@MinSaludCol	144
1200090567985811458	114
@claudiapalacios	104
1144323474808344577	100
1121450757608361984	81
1131955953241350144	58.16666667
1177297241628913664	56
1165653695436140544	55
1111392985126715392	54
1109964801114099714	54
1121395857541271552	54
1121194010020864000	54
1154228871316856834	54
1165984247712276480	54
1176504863641034753	54
1132079823982088192	53

Closeness centrality



For this measure, the result was very disperse considering a high quantity of nodes with different values, making difficult the identification of important entities.

As a summary, this are the more relevant entities per every metric. Additionally colored for the entities that appears two times, three times and four times.

Degree	In-Degree	Out-degree	Betweenness centrality
@MigracionCol	@MigracionCol	@JFernandeznino	@MigracionCol
@JFernandeznino	@petrogustavo	@MigracionCol	@felipemgomez8
@felipemgomez8	@IvanDuque	@felipemgomez8	@ELTIEMPO
@ELTIEMPO	@ELTIEMPO	@aesgx_	@OPSOMS_Col
@aesgx_	@UninorteCO	@JuanNavarrete_	@MinSaludCol
@JuanNavarrete_	@OIMColombia	1154228871316856834	1200090567985811458
@petrogustavo	#Colombia	@ELTIEMPO	@claudiapalacios
1154228871316856834	@lcvelez	1132401199204970496	1144323474808344577
@UninorteCO	#AEstaHora	1132079823982088192	1121450757608361984
@OPSOMS_Col	#migrantes	1187043585885134851	1131955953241350144
1132401199204970496	@BluRadioCo	1210633554666897409	1177297241628913664
1132079823982088192	@lafm	1208950019627372545	1165653695436140544
1187043585885134851	@jguaido	@carlosevilma	1111392985126715392
1210633554666897409	@felipemgomez8	@OPSOMS_Col	1109964801114099714
1208950019627372545	@elespectador	1088875762709643264	1121395857541271552
@MinSaludCol	@NoticiasCaracol	1131991450638573568	1121194010020864000
@IvanDuque	@migravenezuela	1155265790842720256	1154228871316856834
@carlosevilma	#Repost	@MinSaludCol	1165984247712276480
1088875762709643264	@USAID_Colombia	1100763592981401600	1176504863641034753
1131991450638573568	@NicolasMaduro	1131955953241350144	1132079823982088192

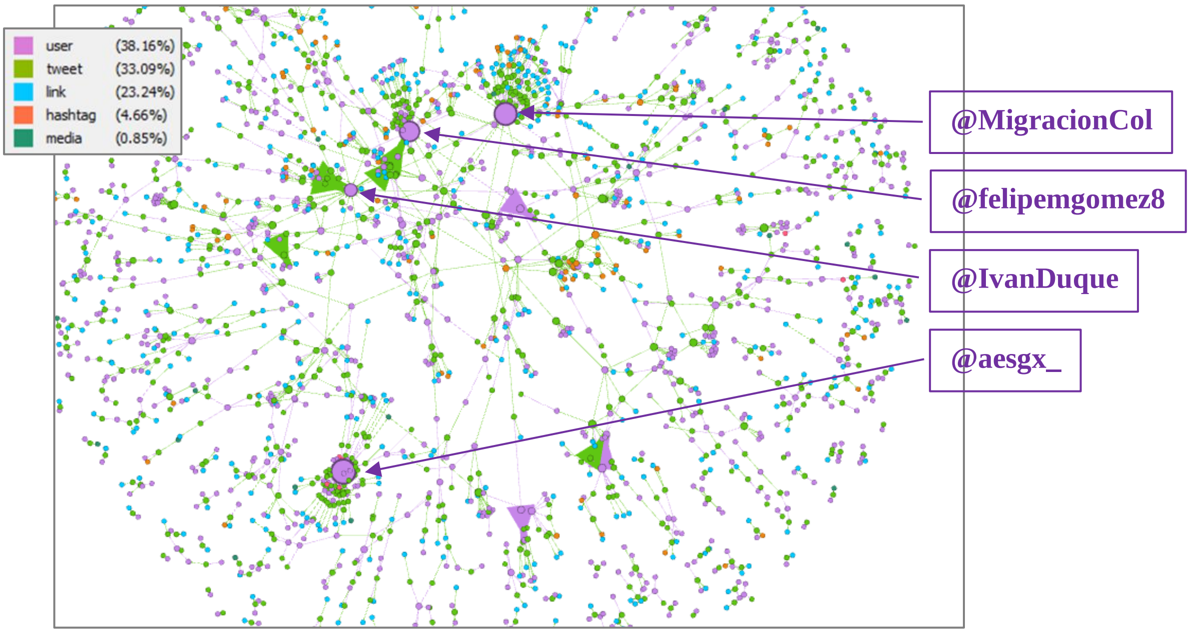
It is important to notice that @MigracionCol, @felipegomez8 and @ELTIEMPO played a huge role in the conversation, with high values in all the metrics. this shows they great relevance, they create and apport content but are also mentioned and included by other entities and they act as a bridge to connect the information. In the same way stands 1154228871316856834, 1132079823982088192 and @MinSaludCol with 3 high values, excluding the in-degree, which means they was not mentioned by other entities.

We can also identify that the out-degree value contributes more to the degree than the in-degree. Making the entities with higher out-degree values also the ones with higher degree values, this means, the entities that creates or contains content.

Lastly, using the Gephi's PageRank methodology we have the list of the ten entities with the higher relevance according to the rank metric:

Degree
@MigracionCol
@petrogustavo
@ELTIEMPO
@IvanDuque
#AEstaHora
@Orianna_vgl
@lcvelez
#Colombia
@Avianca
#QuieroSerPeruanoPorque

f. 2020

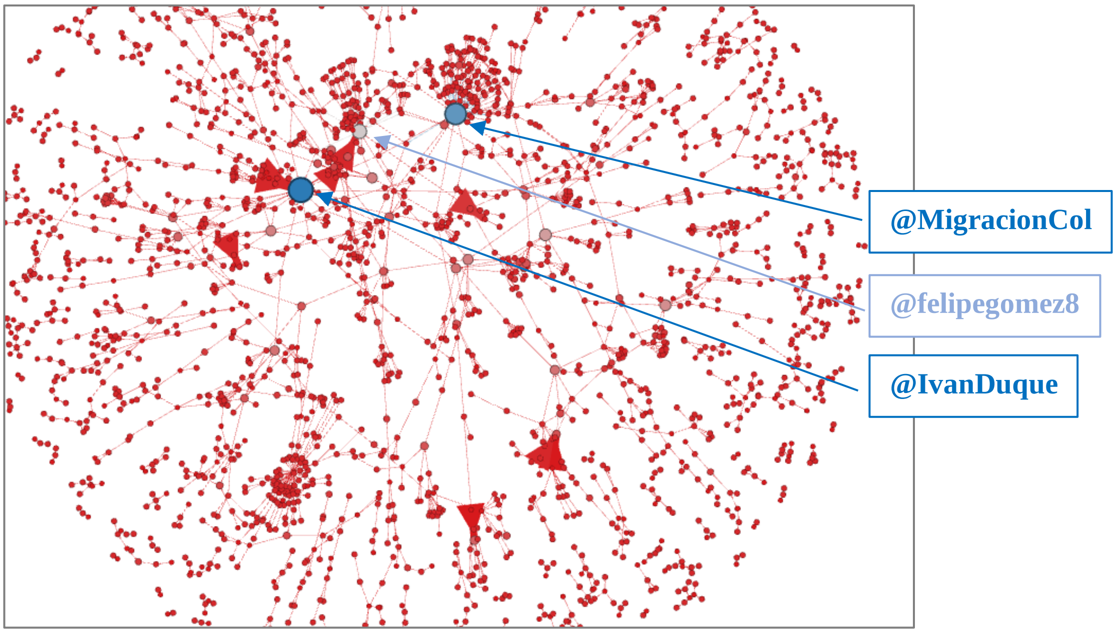


In this visualization can be identified the more relevant entities for the specific year. For example, using the degree we can identify that the users shown above @felipemgomez8, @aesgx_, @MigracionCol, and @IvanDuque) are the most important in terms of the connections they have with other entities. This visualization was made according to the entity type but considering for the size the Degree value. This are the twelve entities with the higher Degree values:

Entity	Degree value
@aesgx_	60
@MigracionCol	54
@felipemgomez8	47
@IvanDuque	25
@Kevinruiz04	11
1276972195512832003	11
1276141514167595008	11
@asangelcolombia	10
1222195942524424192	10
1242958113499688962	10
1255078661746831360	10
1253904485895802880	10
1264686384352288769	10
1296951384898785280	10
1308822591730257922	10
1331554238762283008	10
1331453894816374784	10
#Colombia	9
@JFernandeznino	9
1244023275379884033	9

Additional metrics will be used to have other relevant entities with diverse criteria.

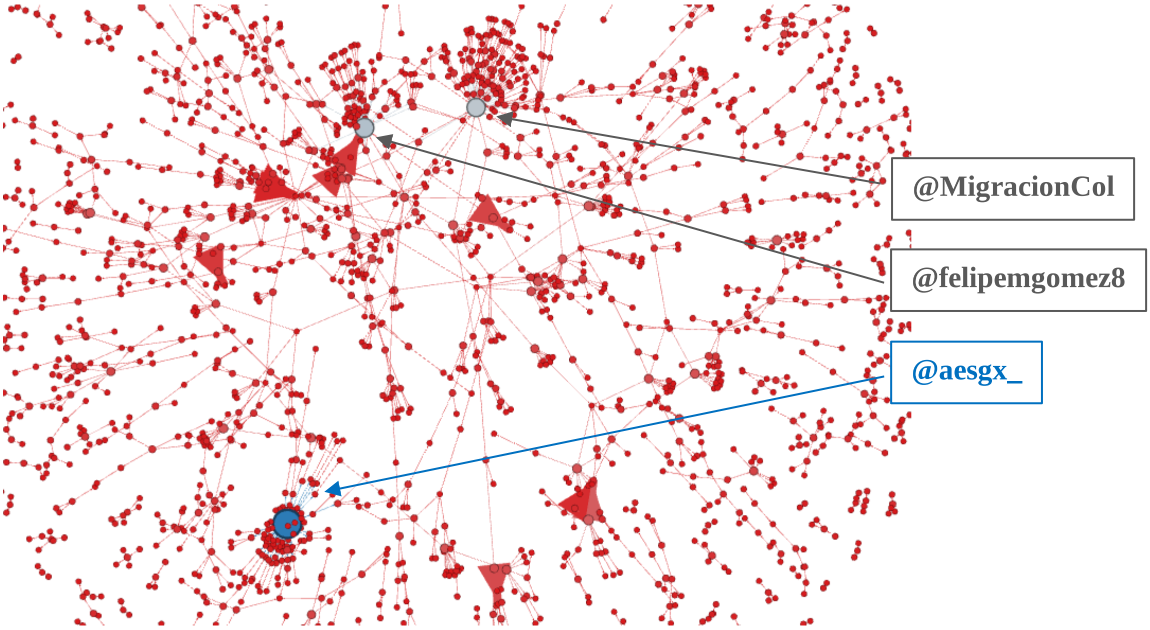
In-Degree



With this metric, we obtain that @IvanDuque has the higher value, this means that in this year most of the people tweet or mention them, gaining importance. There are also other with lower values but also important like @MigracionCol, and @felipemgomez8. This are the twelve entities with the higher In-Degree values:

Entity	In-Degree value
@IvanDuque	25
@MigracionCol	21
@felipemgomez8	12
#Colombia	9
@maisongtv	8
@petrogustavo	7
@ClaudiaLopez	7
@infopresidencia	7
@NicolasMaduro	6
#Venezuela	6
@AlcaldiaCucuta	6
@NoticiasCaracol	5
@AlvaroUribeVel	5
@ACNURamericas	5
#Pasto	5
#AEstaHora	5
@EmmaClaudiaC	5
@SanguinoKenny	4
@ELTIEMPO	4
#Migración	4

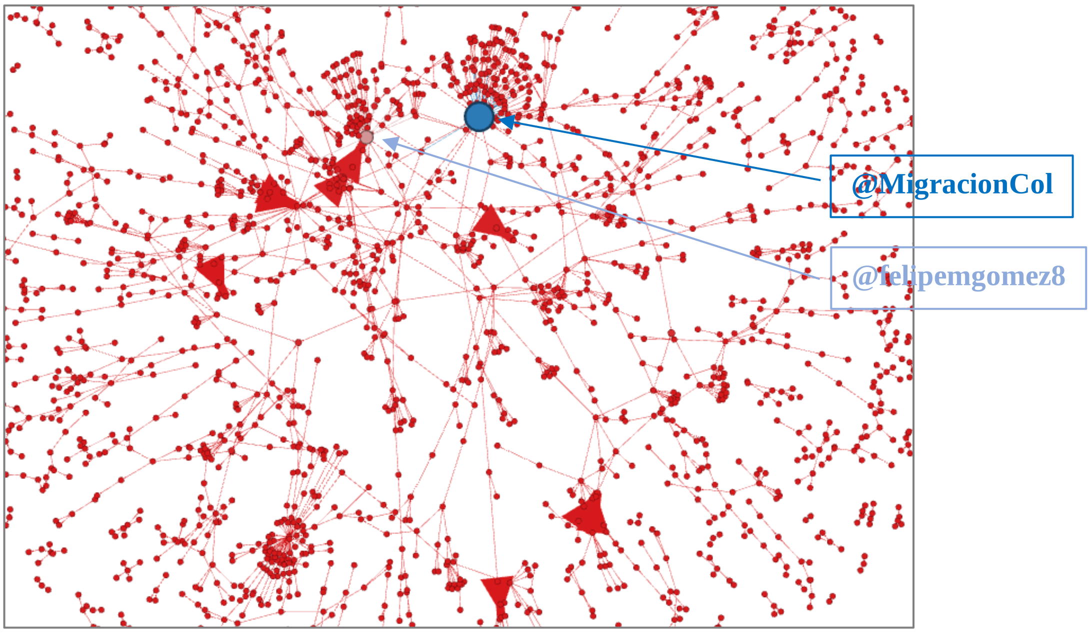
Out-Degree



With this metric, we obtain that @MigracionCol, @aesgx_, and @felipemgomez8 have the higher value, this means that in this year the account tweeted and retweeted very often, being a content generator in the conversation. This are the twelve entities with the higher Out-Degree values:

Entity	Out-Degree value
@aesgx_	59
@felipemgomez8	35
@MigracionCol	33
@Kevinruiz04	11
1276972195512832003	10
1276141514167595008	10
@asangelcolombia	9
1222195942524424192	9
1242958113499688962	9
1255078661746831360	9
1253904485895802880	9
1264686384352288769	9
1296951384898785280	9
1308822591730257922	9
1331554238762283008	9
1331453894816374784	9
1244023275379884033	8
1255277438856237056	8
1296858036284293122	8
1332475216719720449	8

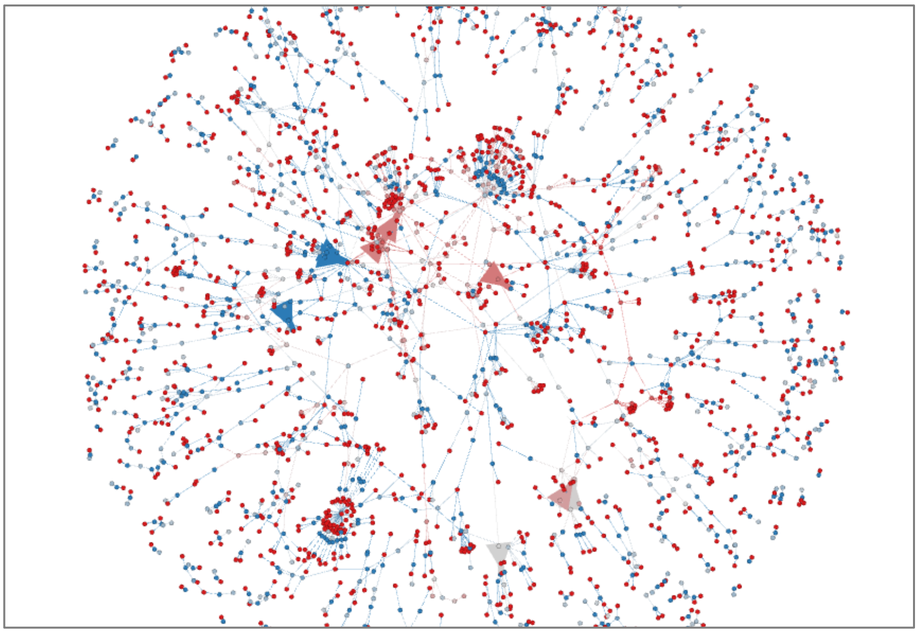
Betweenness centrality



With this metric, we obtain that @MigracionCol has the higher value, this means that in this year the information passed often through this account, being a content connector in the conversation. There are also other with lower values but also important like @felipemgomez8. This are the twelve entities with the higher Betweenness centrality values:

Entity	Betweenness centrality
@MigracionCol	23384
@felipemgomez8	7829.5
@ingjairoyanez	1380
@OIMColombia	945.5
@MinSaludCol	918
@AnaPaolaAgudelo	697
1287938876535963648	684
@ELTIEMPO	666
1287404704645943297	561
1310232980191940613	561
1233216616378380290	483.5
1288133062082072578	478
@JFernandeznino	463
1287751595334291456	436.3
1287094843223814154	436.3
@mluciaramirez	432
1232641791326855168	396
1298310632215904257	374
1310228840111800320	374
1308412809366114309	374

Closeness centrality



For this measure, the result was very disperse considering a high quantity of nodes with different values, making difficult the identification of important entities.

As a summary, this are the more relevant entities per every metric. Additionally colored for the entities that appears two times, three times and four times.

Degree	In-Degree	Out-degree	Betweenness centrality
@aesgx_	@IvanDuque	@aesgx_	@MigracionCol
@MigracionCol	@MigracionCol	@felipemgomez8	@felipemgomez8
@felipemgomez8	@felipemgomez8	@MigracionCol	@ingjairoyanez
@IvanDuque	#Colombia	@Kevinruiz04	@OIMColombia
@Kevinruiz04	@maisongtv	1276972195512832003	@MinSaludCol
1276972195512832003	@petrogustavo	1276141514167595008	@AnaPaolaAgudelo
1276141514167595008	@ClaudiaLopez	@asangelcolombia	1287938876535963648
@asangelcolombia	@infopresidencia	1222195942524424192	@ELTIEMPO
1222195942524424192	@NicolasMaduro	1242958113499688962	1287404704645943297
1242958113499688962	#Venezuela	1255078661746831360	1310232980191940613
1255078661746831360	@AlcaldiaCucuta	1253904485895802880	1233216616378380290
1253904485895802880	@NoticiasCaracol	1264686384352288769	1288133062082072578
1264686384352288769	@AlvaroUribeVel	1296951384898785280	@JFernandeznino
1296951384898785280	@ACNURamericas	1308822591730257922	1287751595334291456
1308822591730257922	#Pasto	1331554238762283008	1287094843223814154
1331554238762283008	#AEstaHora	1331453894816374784	@mluciaramirez
1331453894816374784	@EmmaClaudiaC	1244023275379884033	1232641791326855168
#Colombia	@SanguinoKenny	1255277438856237056	1298310632215904257
@JFernandeznino	@ELTIEMPO	1296858036284293122	1310228840111800320
1244023275379884033	#Migración	1332475216719720449	1308412809366114309

With this, we can identify that the out-degree value contributes more in the degree than the in-degree. Making the entities with higher out-degree values also the ones with higher degree values, this means, the entities that creates or contains content. In this year, most of the entities with high in-degree values are politicians, that reflect how many entities mentions them giving a political environment to the conversation. @MigracionCol and @felipemgomez8 played a huge role in the conversation, with high values in all the metrics. this shows they great relevance, they create and apport content but are also mentioned and included by other entities and they act as a bridge to connect the information.

Lastly, using the Gephi's PageRank methodology we have the list of the ten entities with the higher relevance according to the rank metric:

Degree
@IvanDuque
@MigracionCol
@maisongtv
@felipemgomez8
@petrogustavo
#Colombia
@elespectador
@ClaudiaLopez
@NicolasMaduro
@Citytv

2. Sentiment Analysis and Xenophobia identification

a. Approach 1

The analysis is divided in two parts: a xenophobia classifier and a sentiment analysis. Both work with a lexicon approach, which means a dictionary of words is used to contrast them with every word of the text and determine a metric o score for every tweet.

The lexicon from the first part is based in hate speech words that normally the Colombians use to insult people, many of them are from Colombiamágica[2]. In which, there is a segmentation between insults to women and to men, using the gender of the insulting word. For every tweet it contrasts its words to the lexicon and if at least one word is contained, the tweet is classified with hate speech.

Additionally, the lexicon from the second part was obtained from the afinn R package documentation[3]. It contains a list of words in Spanish and English and a tone score for every of them in a scale from -5 (the most negative) to 5 (the most positive). For every tweet it contrasts its words to the lexicon and calculate the sum of the tone associated to all words, It is the score of the tweet.

---

[2] Los Insultos y groserías más populares de Colombia | Groserías de Colombia (colombiamagica.co)
[3] RPubs - Análisis de sentimientos con R - Léxico Afinn

1. Words Frequency (global)

Word	Frequency
maduro	75
gracias	73
crisis	72
solo	47
apoyo	46
problema	45
mejor	44
ayuda	36
favor	28
mal	27
bueno	26
mayor	26
guerra	25
violencia	24
solidaridad	23
grande	22
oportunidades	22
hambre	21
peor	20

Frequency table of the words

Wordcloud of the most frequent words

As we can see in the figures, we have some words that have a great frequency, also considering that is data for 5 years. Also, there is variability in the tone of the words. For example, *crisis* and *problema* have negative connotation, but *apoyo* and *gracias* no. To identify better these words that drove the global tone, will be created an additional graph that contains this words with a different color based on the tone.

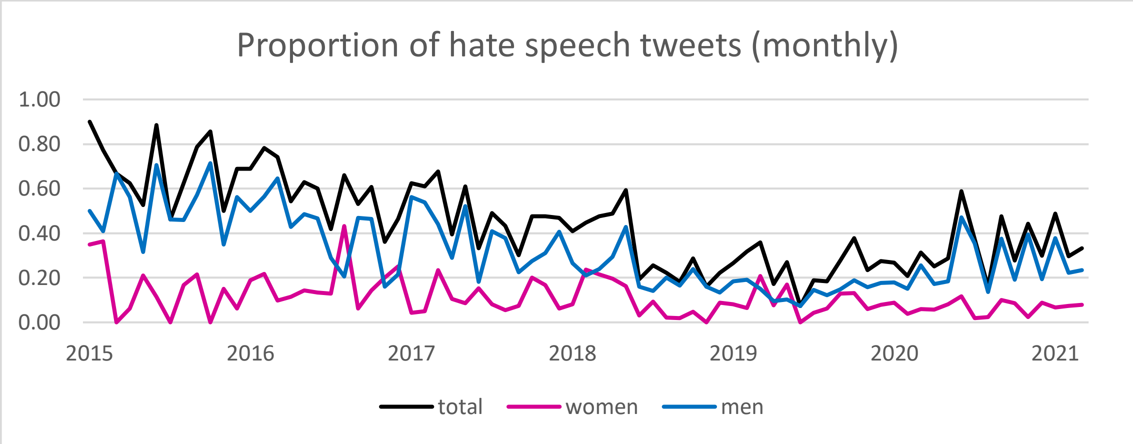
2. Tone Frequency (global)

Tone	Abs. Frequency	Rel. Frequency	Gruop Frequency
-15	1	0.03%	25.25%
-11	1	0.03%	
-10	2	0.07%	
-9	6	0.20%	
-8	9	0.30%	
-7	9	0.30%	
-6	24	0.80%	
-5	50	1.67%	
-4	77	2.58%	
-3	144	4.82%	
-2	291	9.73%	
-1	141	4.72%	
0	1574	52.64%	52.64%
1	141	4.72%	22.11%
2	290	9.70%	
3	117	3.91%	
4	51	1.71%	
5	32	1.07%	
6	15	0.50%	
7	5	0.17%	
8	6	0.20%	
9	2	0.07%	
10	1	0.03%	
11	1	0.03%	

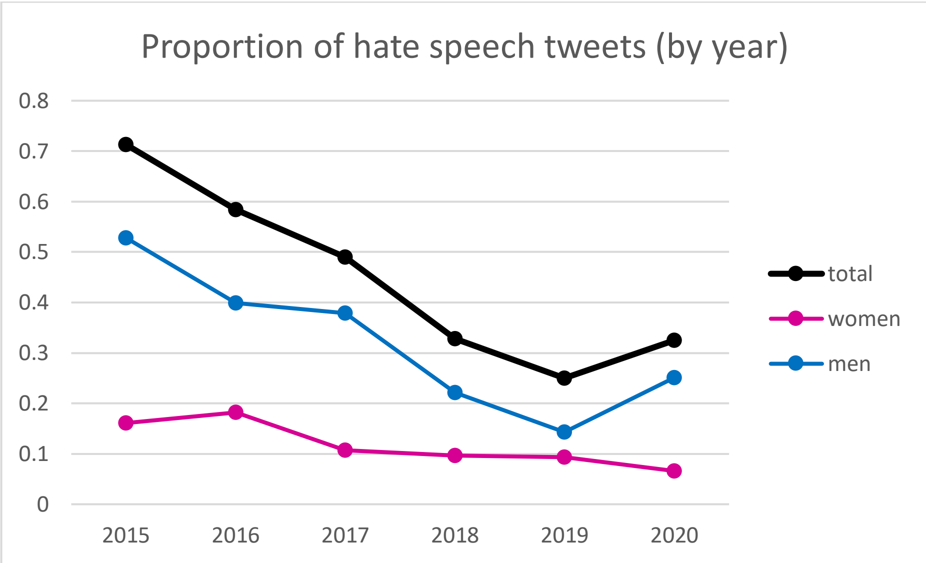
Frequency table of the tones (globally)

This table show us that although having a neutral tone is the most frequent situation in the tweets, between positive and negative tones the negative ones are more common. Also, we can identify that it's not so frequent to find high negative or high positive tweets, they are mostly between the -5 and 5 interval. An additional approach is in the next figures, that shows the general evolution of the average tone through the years.

4. Time series from the Xenophobia classifier:



Time series of the proportion of hate speech tweets globally and monthly.



Time series of the proportion of hate speech tweets globally and by gender yearly.

As we can see, the proportion of hate speech tweets is decreasing through time (more than 50% from 2015 to 2019), this complements the past figure which shows that also the tone is becoming more positive. Additionally, there is a constant hate speech tweets proportion to women in all the years, but to men (same as the global proportion) is decreasing (around 66%). So, the people is becoming more polite talking about the men, but the negative connotation to women is mostly the same.

5. Words that conveyed xenophobia (global)

The words presented below were the ones more commonly used in the pejorative talking toward the migrants. Considering that in Spanish there are different articles and adjectives depending on the subject, in the dataset was included for every insult a different version for masculine and feminine, but also for singular and plural.

Men	
Word	Frequency
veneco	864
venecos	17
perro	6
loco	4
maldito	4
pendejos	3
perros	3
malparido	3
pendejo	3
maricón	2
flojo	2
bobos	2
mamertos	2
cachon	2
maricon	1
puto	1
corroncho	1
vagos	1

Frequency table of the xenophobic words to men

Women	
Word	Frequency
veneca	322
puta	7
putas	5
prostitutas	3
maldita	3
venecas	3
loca	3
locas	2
perra	2
insornia	1
boba	1
guisa	1
pendeja	1
bobas	1

Frequency table of the xenophobic words to women

Nongender	
Word	Frequency
hijueputa	12
jueputa	5
idiotas	3
idiota	2
hijueputas	2
guaricha	1
gonorrea	1
huevas	1
imbécil	1
imbecil	1

Frequency table of The non genderized xenophobic words

Wordcloud of the most frequent xenophobic words to men

Wordcloud of the most frequent xenophobic words to men

Wordcloud of the most frequent xenophobic words to men

However, if we want to synthesize the results combining words with the same grammatical root (omitting the plurality and spelling errors) we have the following:

Men	
Word	Frequency
veneco	881

Women	
Word	Frequency
veneca	325

Nongender	
Word	Frequency
hijueputa	19

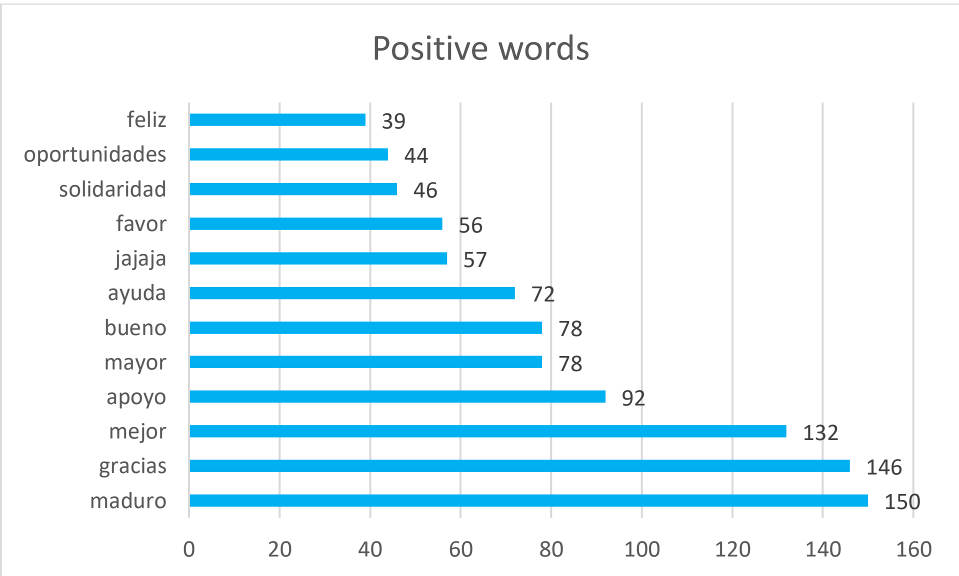
perro	9
pendejo	6
loco	4
maldito	4
malparido	3
maricón	3
flojo	2
bobos	2
mamerto	2
cachon	2
puto	2
guevon	2
ñero	2
vago	2
corroncho	1
pirobo	1
sapo	1
petardo	1

puta	12
loca	5
prostituta	3
maldita	3
perra	2
boba	2
insornia	1
guisa	1
pendeja	1

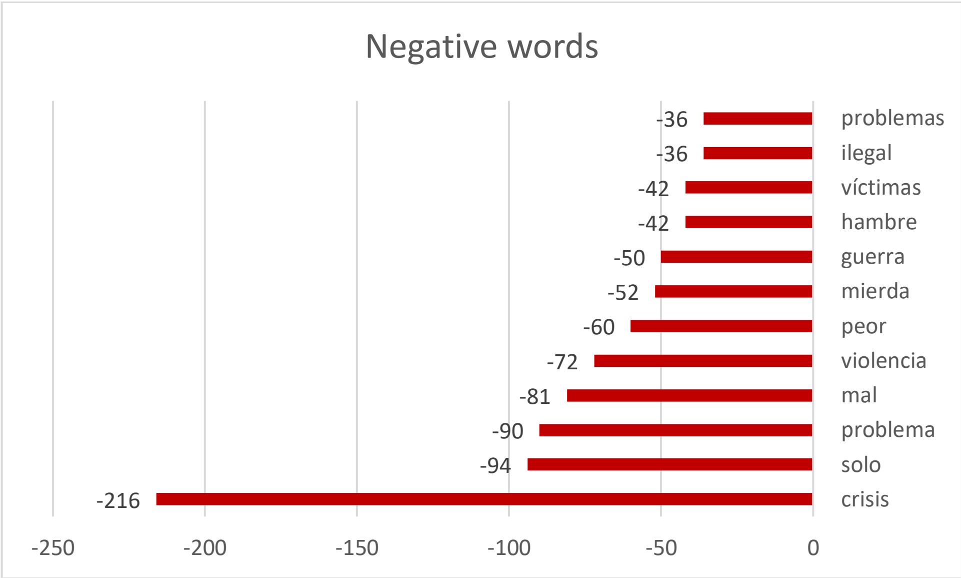
idiota	5
imbécil	2
guaricha	1
gonorrea	1
hueva	1

As we can see in the tables and the figures, the most dominant hate speech words used were *veneco* and *veneca*, a slang commonly used by Colombians referring to Venezuelans (in a pejorative way). Also, some other insults were also present like *hijueputa* or *perro*, but in general the proportion is not so big.

6. Words that moved the tone (global)



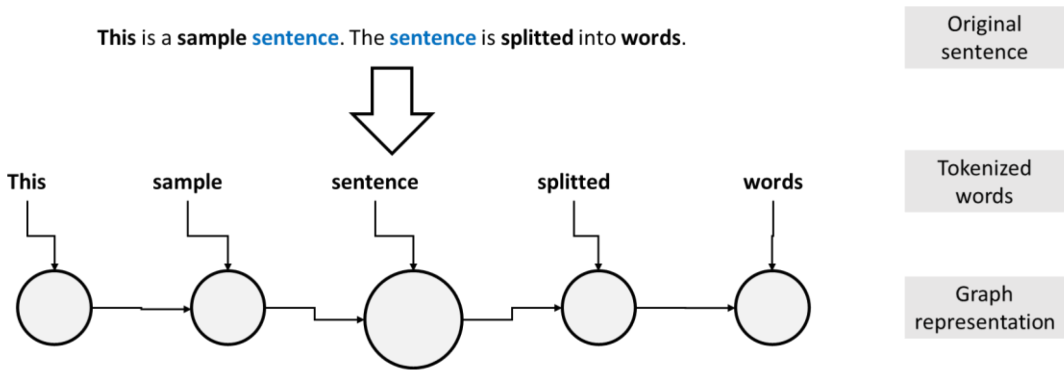
Barplot of the positive words with more weight.



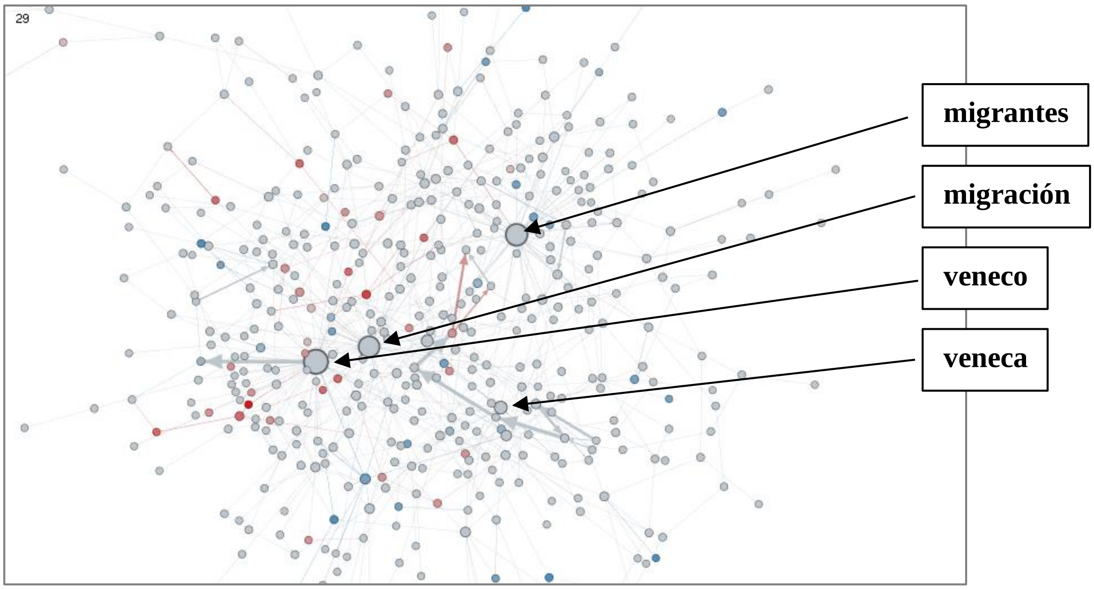
Barplot of the negative words with more weight.

b. *Text analysis (from the co-occurrence Graph)*

The co-ocurrence graph is used to represent clearly large amounts of texts. This methodology uses a tokenization of a sentence into words and connects them as they are near. So, the connections represent when two words are together, and the directions goes from the first word to the second one. With this, we can see which words were often used near other ones. Those two words form a bigram. This process is represented below:



1. **2015**



The conversation in this year was mostly around the words above (migrantes, migración, veneco and veneca). Additionally, below are the ten words with higher degree (frequency), and the ten connections with more weight (frequency). As *migrantes* had the higher degree value, this suggests that it was focused mainly on the migrant people, with a negative connotation due the degree value of *veneco* and *veneca*.

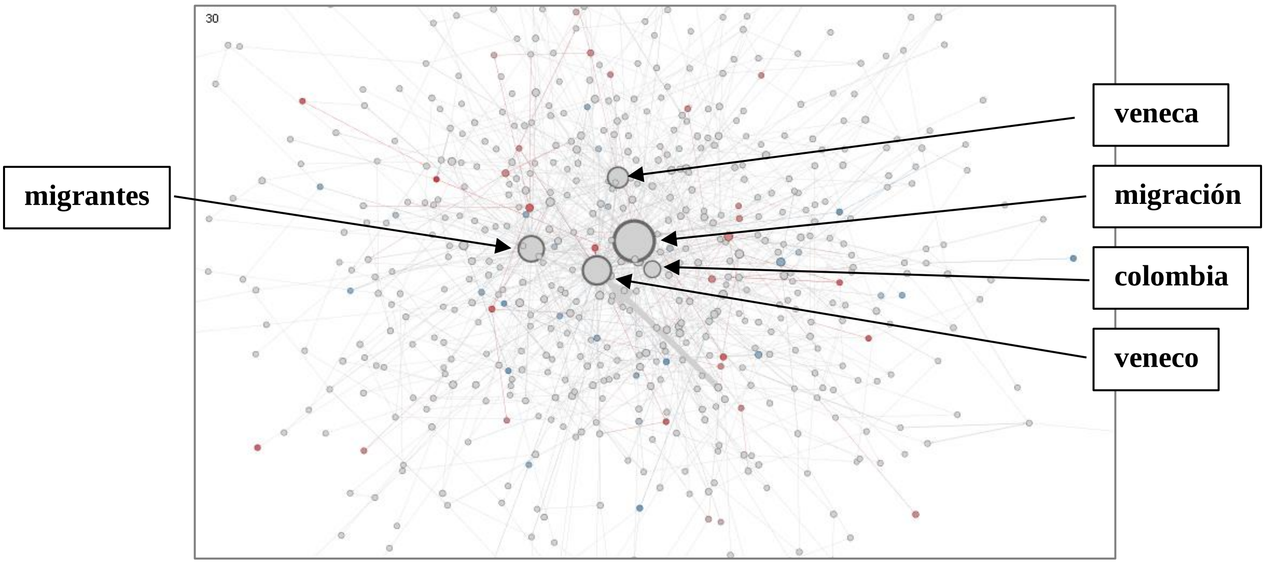
Word	Frequency
veneco	176
migrantes	136
migración	124
veneca	65
bogotá	32
colombia	32
venezuela	23
maduro	22
debe	19
colombianos	16

Edge		Frequency
Source	Target	
valen	huevo	26
vale	lucha	26
veneco	unido	26
huevo	vale	24
declaraciones	países	19
lucha	pueblo	16
países	valen	14
flor	día	11
persistente	pueblo	10
cumbre	iberoamericana	9

The first table shows the key words (mostly used with other ones) in the tweets, and the second table the most frequent bigrams.

Additionally, we have a wordcloud with the most frequent words used with *veneco* and *veneca*, respectively.

2. **2016**



The conversation in this year was mostly around the words above (migrantes, migración, veneco, veneca and Colombia). Additionally, below are the ten words with higher degree (frequency), and the ten connections with more weight (frequency). As *migración* had the higher degree value, this suggests that it was focused mainly on the migration situation and less on the people, The negative connotation remains due the degree value of *veneco* and *veneca*.

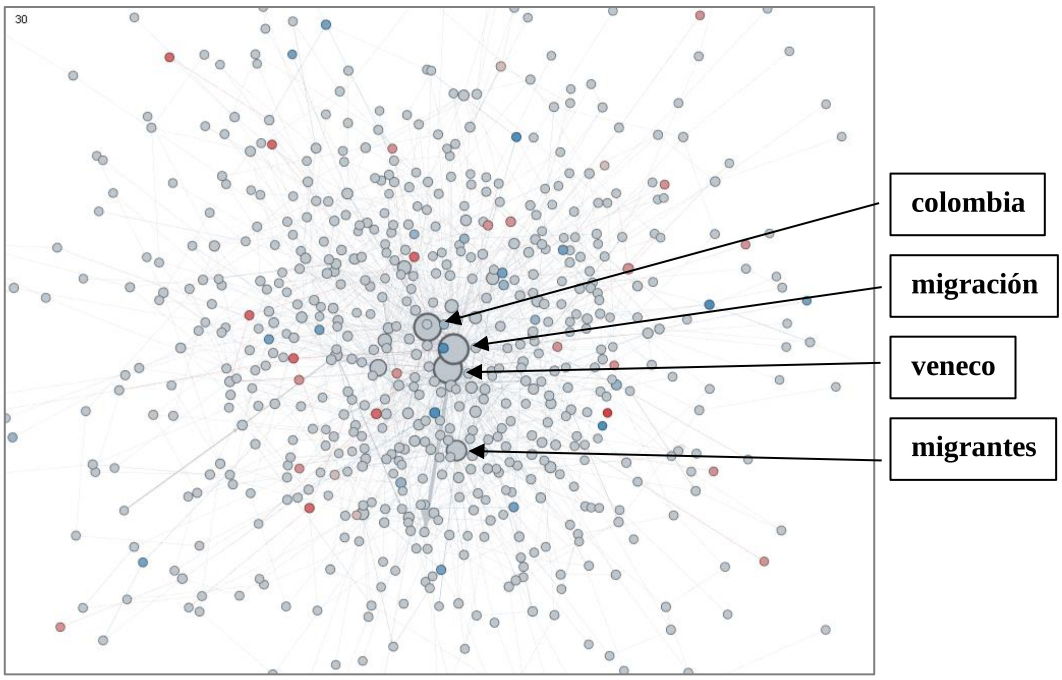
Word	Frequency
migración	236
veneco	156
migrantes	138
veneca	105
colombia	77
migracion	35
hoy	24
tráfico	22
solo	22
frontera	21

Edge		Frequency
Source	Target	
pueblo	veneco	33
firma	electrónica	6
hijo	puta	5
traficantes	migrantes	4
colombia	arauca	4
migrantes	puede	4
golazo	veneco	4
colombia	regional	4
veneco	vamos	3
migrantes	centroamericanos	3

The first table shows the key words (mostly used with other ones) in the tweets, and the second table the most frequent bigrams.

Additionally, we have a wordcloud with the most frequent words used with *veneco* and *veneca*, respectively.

3. 2017



The conversation in this year was mostly around the words above (migrantes, migración, veneco, and Colombia). Additionally, below are the ten words with higher degree (frequency), and the ten connections with more weight (frequency). As *migración* had the higher degree value, this suggests that it was focused mainly on the migration situation and less on the people, The negative connotation remains due the degree value of *veneco* and *veneca*.

Word	Frequency
migración	254
veneco	210
colombia	140

Edge		Frequency
Source	Target	
tráfico	migrantes	37
migrantes	trata	10

migrantes	112
veneca	70
venezolanos	61
país	44
venezuela	40
ser	28
colombianos	26

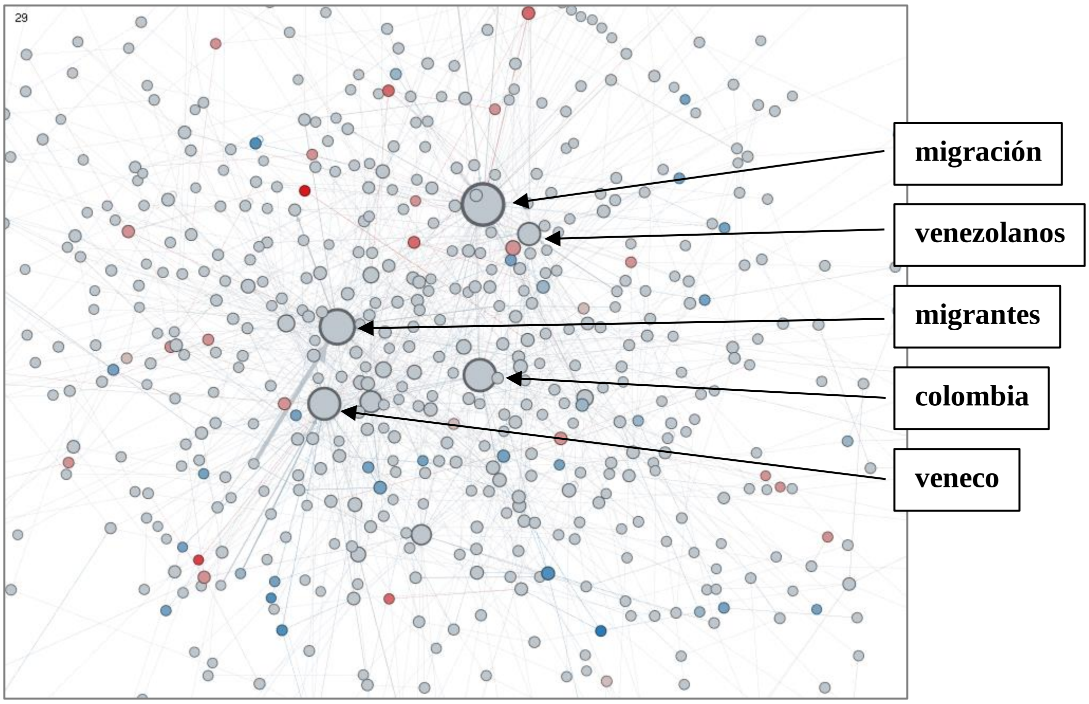
grande	veneco	9
régimen	veneco	9
pasar	migración	8
rescata	migrantes	8
hermanos	venezolanos	8
bogotá	colombia	7
migración	hacia	6
equipo	veneco	6

The first table shows the key words (mostly used with other ones) in the tweets, and the second table the most frequent bigrams.

Additionally, we have a wordcloud with the most frequent words used with *veneco* and *veneca*, respectively.

The phrase "comer veneca" is used refering to have sex with venezuelan prostitutes.

4. 2018



The conversation in this year was mostly around the words above (migrantes, migración, veneco, venezolanos and Colombia). Additionally, below are the ten words with higher degree (frequency), and the ten connections with more weight (frequency). As *migración* had the higher degree value, this suggests that it was focused mainly on the migration situation and less on the people, The negative connotation remains due the degree value of *veneco* and *veneca*.

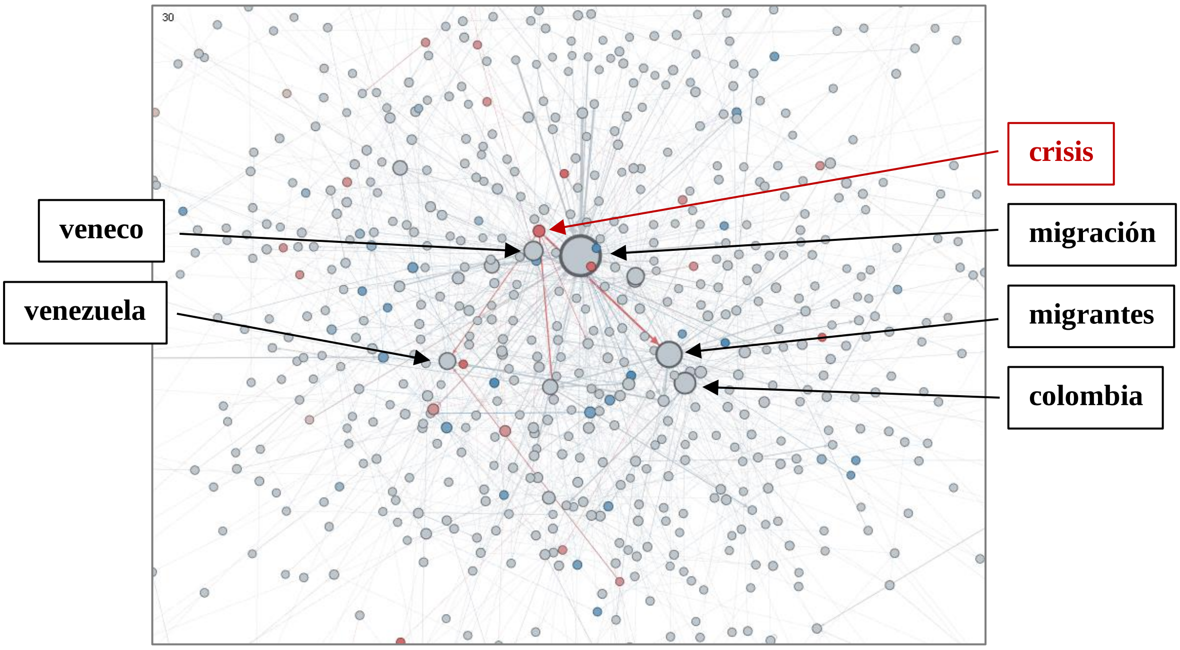
Word	Frequency
migración	170
migrantes	145
veneco	133
colombia	109
venezolanos	63
veneca	55
país	52
venezuela	49
hoy	30
venezolana	30

Edge		Frequency
Source	Target	
migración	canadá	38
aeropuerto	internacional	10
fans	farc	10
farc	ahora	10
ahora	quieren	9
internacional	dorado	7
vamos	veneco	6
gob	veneco	6
desplazados	migrantes	6
personas	tráfico	5

The first table shows the key words (mostly used with other ones) in the tweets, and the second table the most frequent bigrams.

Additionally, we have a wordcloud with the most frequent words used with *veneco* and *veneca*, respectively.

5. 2019



The conversation in this year was mostly around the words above (migrantes, migración, veneco, Venezuela and Colombia). An additional node is *crisis*, it has a negative tone and a considerable number of connections with other words including *migrantes, salud* and *Venezuela*. Additionally, below are the ten words with higher degree (frequency), and the ten connections with more weight (frequency). As *migración* had the higher degree value, this suggests that it was focused mainly on the migration situation and less on the people, The negative connotation remains due the degree value of *veneco* and *veneca*.

Word	Frequency
migración	127
migrantes	70
colombia	53
veneco	44
venezolanos	37
venezuela	35
salud	29
país	28

Edge		Frequency
Source	Target	
foro	migración	16
fila	migración	11
crisis	migrantes	11
oficina	migración	10
colombianos	venezuela	9
migrantes	colombianos	9
migración	mixta	8

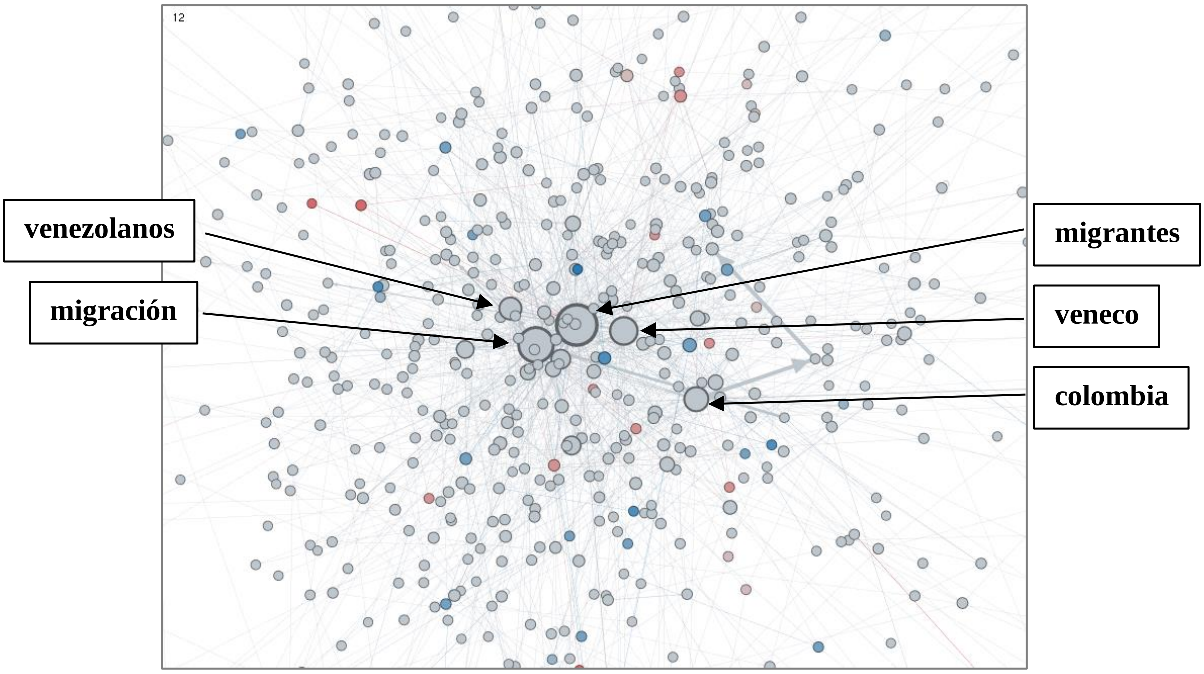
hoy	27
veneca	26

migración	colombianos	8
venezolano	veneco	7
atención	salud	7

The first table shows the key words (mostly used with other ones) in the tweets, and the second table the most frequent bigrams.

Additionally, we have a wordcloud with the most frequent words used with *veneco* and *veneca*, respectively.

6. 2020



The conversation in this year was mostly around the words above (migrantes, migración, veneco, Venezuela and Colombia). Additionally, below are the ten words with higher degree (frequency), and the ten connections with more weight (frequency). As *migrantes* had the higher degree value, this suggests that it was focused mainly on the migrant people, with a negative connotation due the degree value of *veneco* and *veneca*.

Word	Frequency

Edge		Frequency

migrantes	111
migración	91
veneco	65
colombia	53
venezolanos	46
venezuela	35
veneca	33
país	29
ser	22
hoy	21

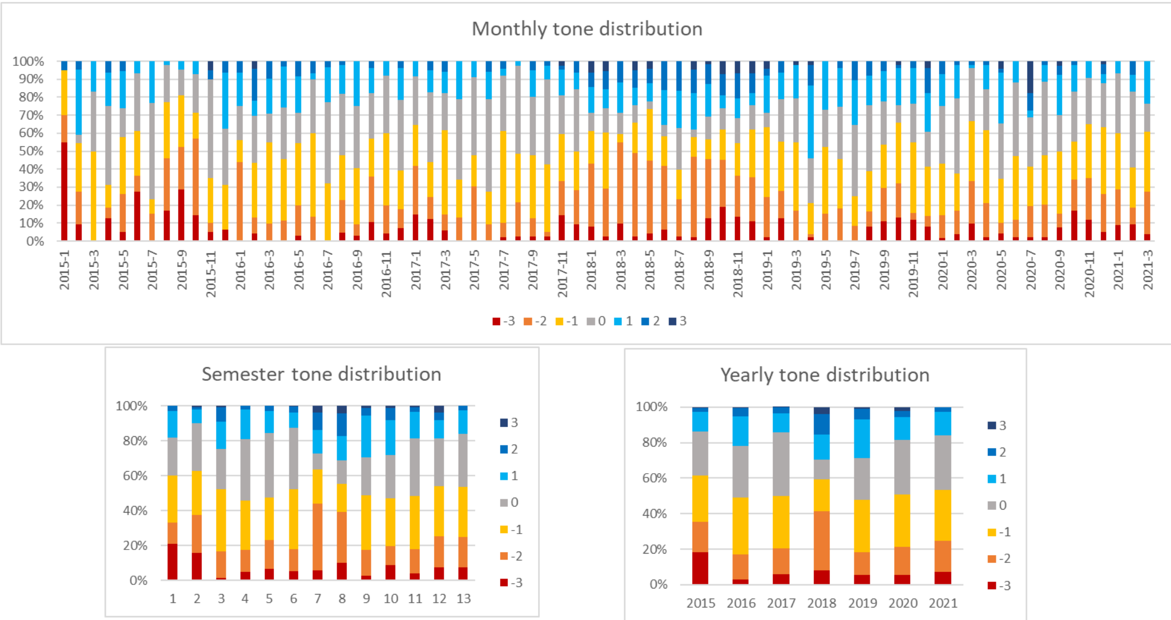
Source	Target	
in	bogotá	71
colombia	in	70
at	migración	48
migración	venezolanos	32
migración	masiva	23
miles	migrantes	16
migrantes	ilegales	15
venezolanos	migrantes	15
niños	niñas	14
migración	irregular	13

The first table shows the key words (mostly used with other ones) in the tweets, and the second table the most frequent bigrams.

Additionally, we have a wordcloud with the most frequent words used with *veneco* and *veneca*, respectively.

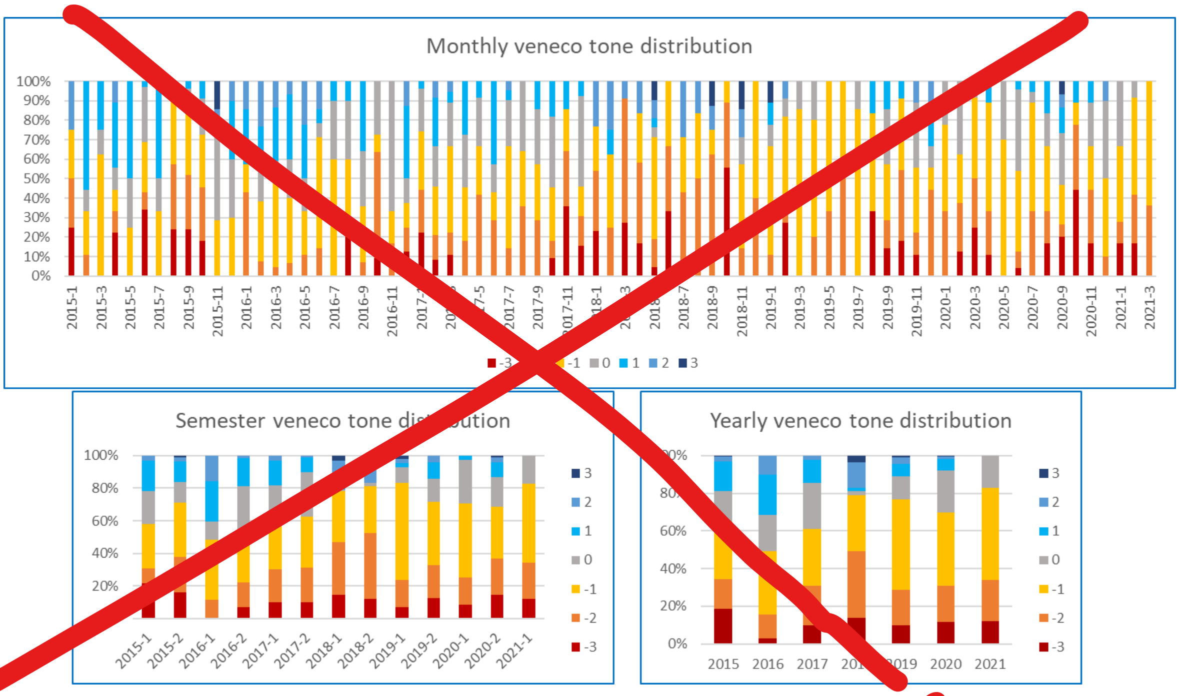
c. *Approach 2 (Manual labeling)*

With the 2990 valid tweets, a manual characterization was made of the tone of the tweet, the tone of the word veneco or veneca (if it was used), and what the tweet mentions. For the tones, a scale from -3 (very negative) to 3 (very positive) was used, taking 0 as neutral. For the actor in question, it was divided into Migrants, Migration, Government (Colombian), Geopolitics (foreign governments), Media, Local and Others. With this classification, a graphic analysis of the results was made.

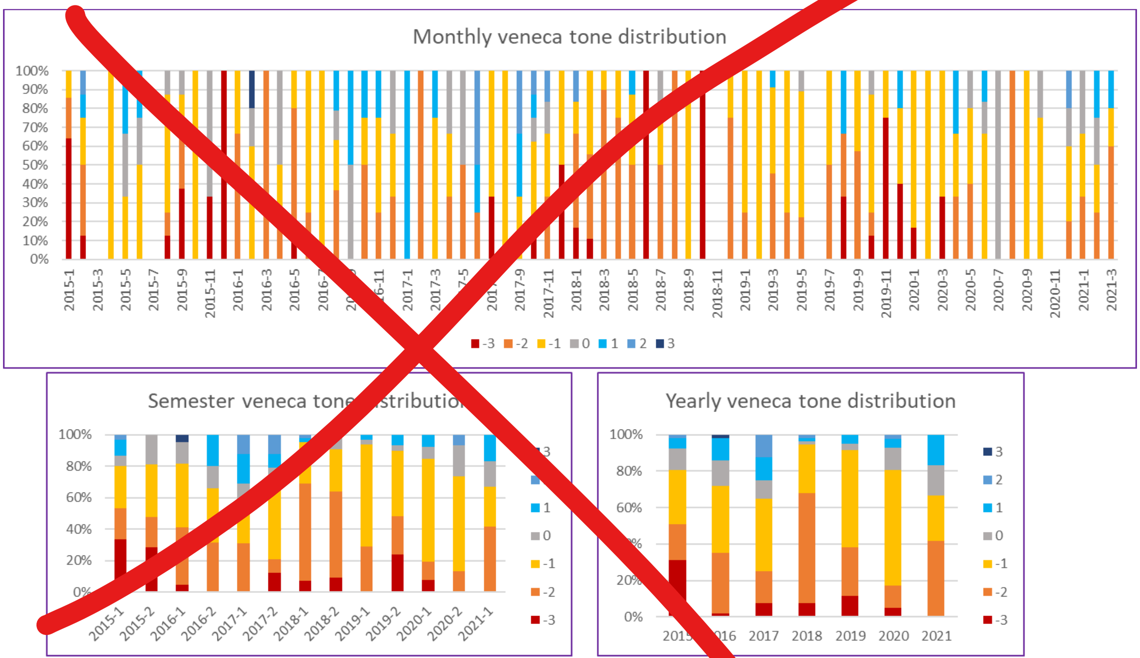


Evolution of the tone distribution.

First, the proportion of tweets with a negative tone towards the phenomenon is similar across years and semesters, although with some fluctuations per month. This shows that the negative perception is largely maintained over the years. The same happens with the proportion of neutral and positive tweets except for 2018 where an increase in the proportion of positive ones was evidenced. 2018 also has the highest proportion of tweets between -3 and -2, showing that in that year (where the phenomenon worsened) many opinions in favor of migrants were generated, but there was also an increase in opinions against.

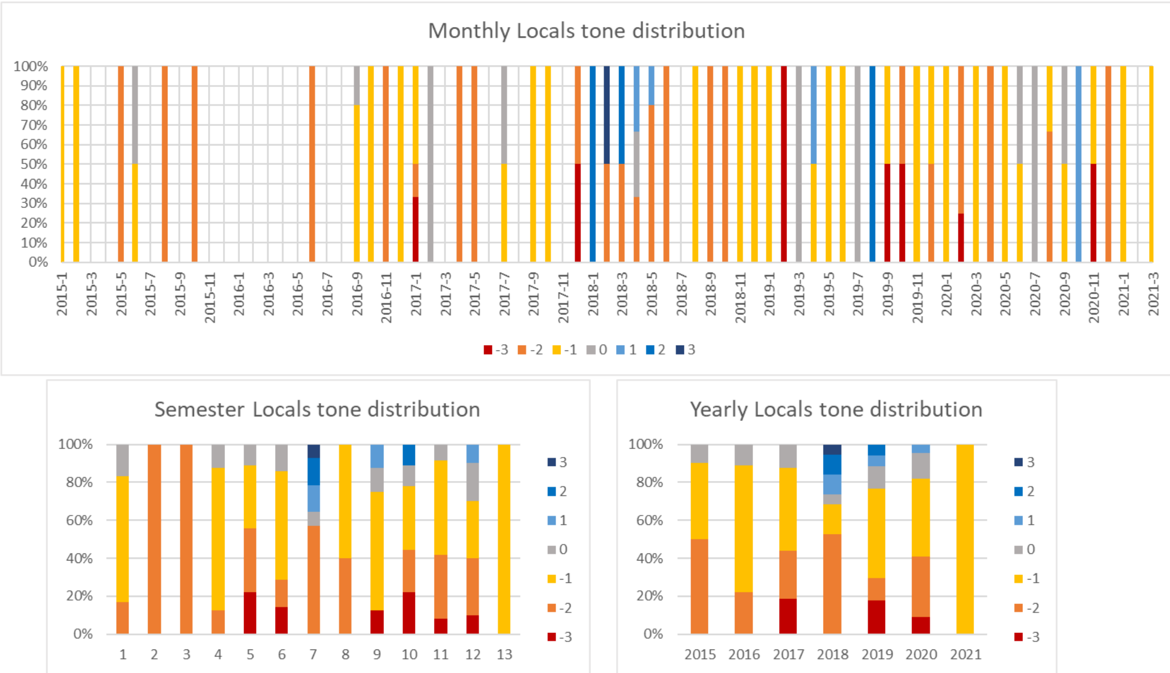


Evolution of the tone distribution of the word veneco



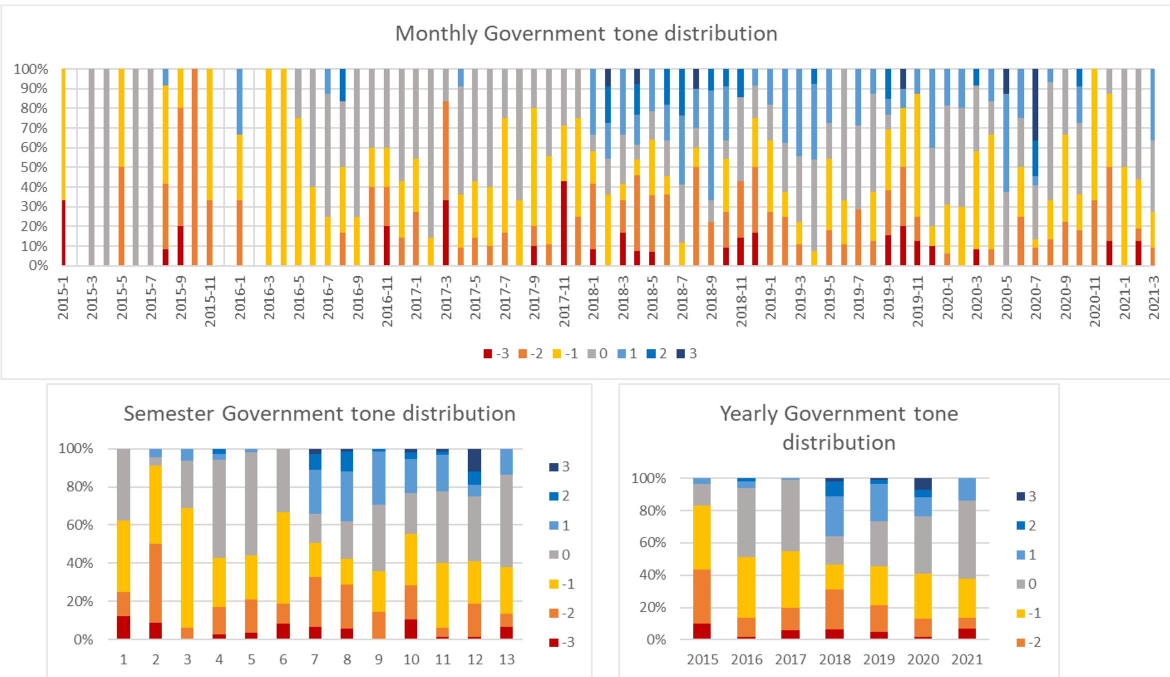
Evolution of the tone distribution of the word veneca

Third, there is a higher proportion of positive tones for veneco than for veneca. This indicates that this term was much more used to refer positively to a Venezuelan migrant than to a Venezuelan migrant. In addition, the proportion of negative tones was much higher for the word veneca, so there is a notable difference in the perception that Colombians have of migrants, which is much more negative for female migrants than for male migrants. Here a gender gap is evident, and how migration can be harder for women migrants than for men migrants given a greater rejection by the locals towards them.



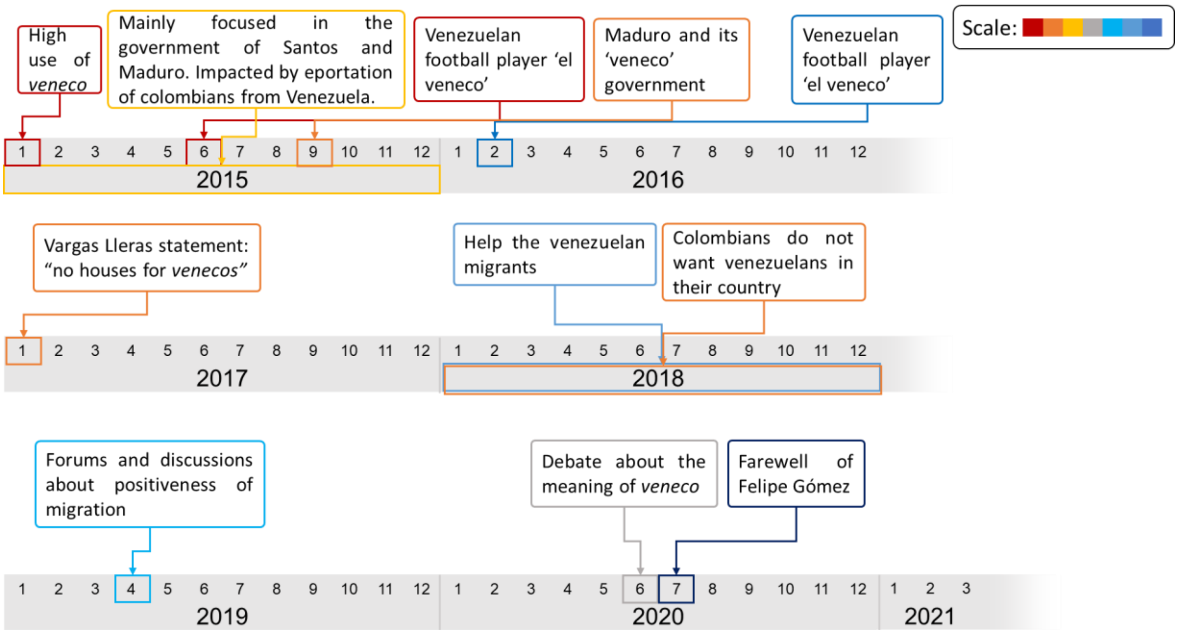
Evolution of the tone distribution toward locals

In fourth place, considering only the tweets that are directed to Colombians, it is evident the high proportion of negative tweets, showing that around the migratory phenomenon the negative perception is not only high towards migrants, but also towards Colombians themselves. As an exception we have 2018, where the aggravation of the phenomenon also encouraged an increase in the proportion of positive tweets.



Evolution of the tone distribution toward the Colombian government

In fifth place, the proportion of negative tweets oscillates around 50% for all years, showing a high negative perception towards the government. However, by 2015 this proportion reached more than 80%, a year in which the deportation of Colombians created tensions between the governments of Venezuela and Colombia, and between Colombians and their rulers (mainly the president of that year, Juan Manuel Santos). Additionally, the proportion of positive tweets was not significant until 2018, showing that there were not many pro-government comments in the first years, these started when the crisis worsened, and the government created policies and aid for migrants.



Timeline of the relevant events

Finally, we have this timeline (in English) whose color represents the -3 to 3 scale previously shown. Here we see that in 2015 and 2016 soccer significantly impacted the tone of the conversation, being very negative in June 2015 and very positive in February 2016, both influenced by the player 'el veneco' (Seijas) but with totally opposite comments. In 2015 the conversation was very related to the Colombian and Venezuelan governments, this was generated from the increase in deportations of Colombians by Venezuela, creating comments of rejection and hatred. increasing the negativity of the conversation in that year. We also have that in 2017 Vargas Lleras' comment "there are no houses for Venezuelans" influenced the proportion of negative comments towards migrants. In 2018 there was a division between opinions, where one part supported migrants arriving in Colombia, but the other rejected them. Contrasting with Figure 1, the proportion is not equal, as the proportion of negative tones was 30% positive and 60% negative. In 2019, an increase in the number of forums and discussions about migrants and the benefits of their inclusion impacted an increase in positive tweets. Finally, in 2020, the farewell of civil servant Felipe Gómez generated a wave of positive tweets due to the appreciation of his work and support for Venezuelan migrants.