

JETTY KOMRIJ
DECEMBER 2020

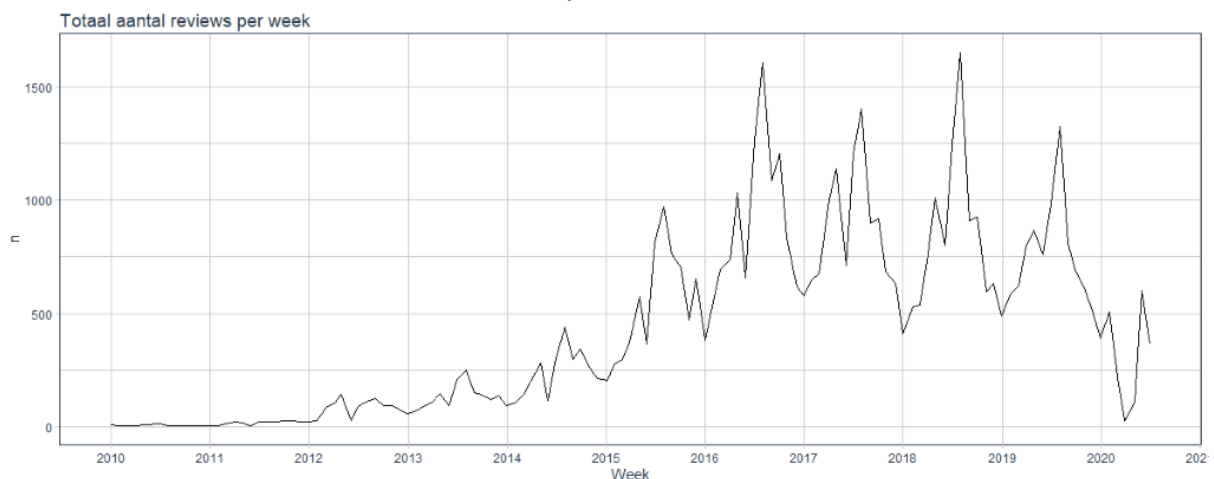
1. Introductie

Waar gaan toeristen naar toe, in welke hotels overnachten toeristen en welke toeristische hotspots bezoeken toeristen gedurende hun reis door Friesland. Om een beeld te krijgen van het bezoekersgedrag van toeristen binnen Friesland is er data verzameld door middel van webscraping, waarbij op een geautomatiseerde manier reviews van een website kunnen worden opgehaald. Voor deze analyse zijn alle bestemmingen in Friesland en bijhorende reviews op de website TripAdvisor gescraped. TripAdvisor is een veelgebruikt website voor het geven van reviews over accommodaties over de hele wereld. TripAdvisor beslaat 10 miljoen reizigers reviews en meningen van 5 miljoen geregistreerde members en wordt maandelijks gebruikt door 25 miljoen bezoekers (Miguéns, Baggio, & Costa, 2008).

Op basis van deze gegevens is een Sociale Netwerk Analyse uitgevoerd. Door de data te analyseren als een netwerk, worden relaties tussen hotels, restaurants en activiteiten inzichtelijk die gecreëerd zijn door toeristen zelf. Door middel van deze benadering kun je daarnaast thematische en geografische groepen identificeren binnen Friesland (Hernández, Kirilenko, & Steochenkova, 2018). Op basis van de geschreven reviews kan inzichtelijk gemaakt worden welke accommodaties achtereenvolgend door toeristen worden bezocht. Hoe deze attracties met elkaar zijn verbonden wordt inzichtelijk gemaakt door middel van een netwerk van linkjes. Deze linkjes zijn gecreëerd door toeristen die deze attracties bezoeken en reviewen en een Sociale Netwerk Analyse laat dit zien. In deze beschrijving wordt uiteengezet wat sociale netwerk analyse is, hoe het in te zetten is voor bovenstaand doeleinde en welke methodische stappen zijn gezet om van ruwe data tot informatie te komen.

1. Materiaal en Methode

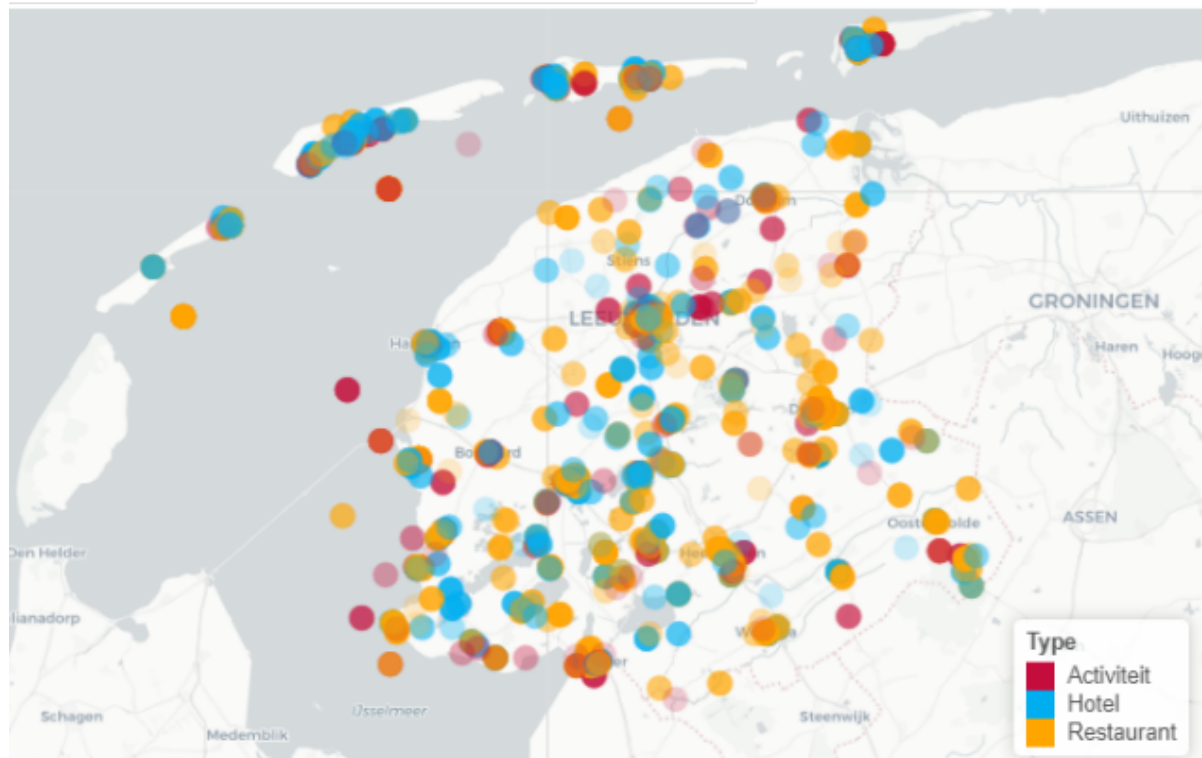
Binnen de sociale netwerk analyse wordt gekeken naar relaties tussen activiteiten. Een relatie tussen accommodaties wordt gevormd als een reviewer meerdere accommodaties in Friesland bezoekt en deze plaatsen verbindt door het geven van een review. Elke reviewer beschikt over een uniek ID. Door middel van de 'rvest' package in R, zijn meer dan 50.000 reviews gescraped over restaurants, hotels en activiteiten in Friesland van 2010 tot en met juli 2020.



Figuur 2: Aantal reviews over tijd accommodaties Friesland Tripadvisor

Pagina-einde

Figuur 1: Accommodaties Fryslân Tripadvisor



De dataset bestaat uit:

- 722 restaurants
 - 33.993 reviews na cleaning
 - 15.928 reviewers
- 317 hotels
 - 15.611 reviews na cleaning
 - 13.470 reviewers
- 216 activiteiten
 - 6.477 reviews na cleaning
 - 3.673 reviewers

Vervolgens is de dataset opgesplitst naar herkomst om meer inzicht te krijgen in het bezoek per herkomst, waarbij onderscheid is gemaakt in:

- Lokale toerist
 - Toeristen die uit Friesland komen en Friesland bezoeken. Deze subset bestaat uit 2.561 gebruikers en 8.436 reviews.
- Nederlandse toerist
 - Toeristen die uit de overige provincies van Nederland afkomstig zijn en Friesland bezoeken. Deze subset bestaat uit 14.174 gebruikers en 30.108 reviews.
- Buitenlandse toerist
 - Toeristen die Friesland bezoeken en niet uit Nederland komen. Deze subset bestaat uit 3.568 gebruikers en 6.122 reviews.
- Toerist
 - Daarnaast wordt het gehele netwerk in kaart gebracht van zowel toeristen uit Friesland, Nederland, buitenland en onbekende herkomst. Het gehele netwerk bestaat uit 27.254 gebruikers en 56.081 reviews.

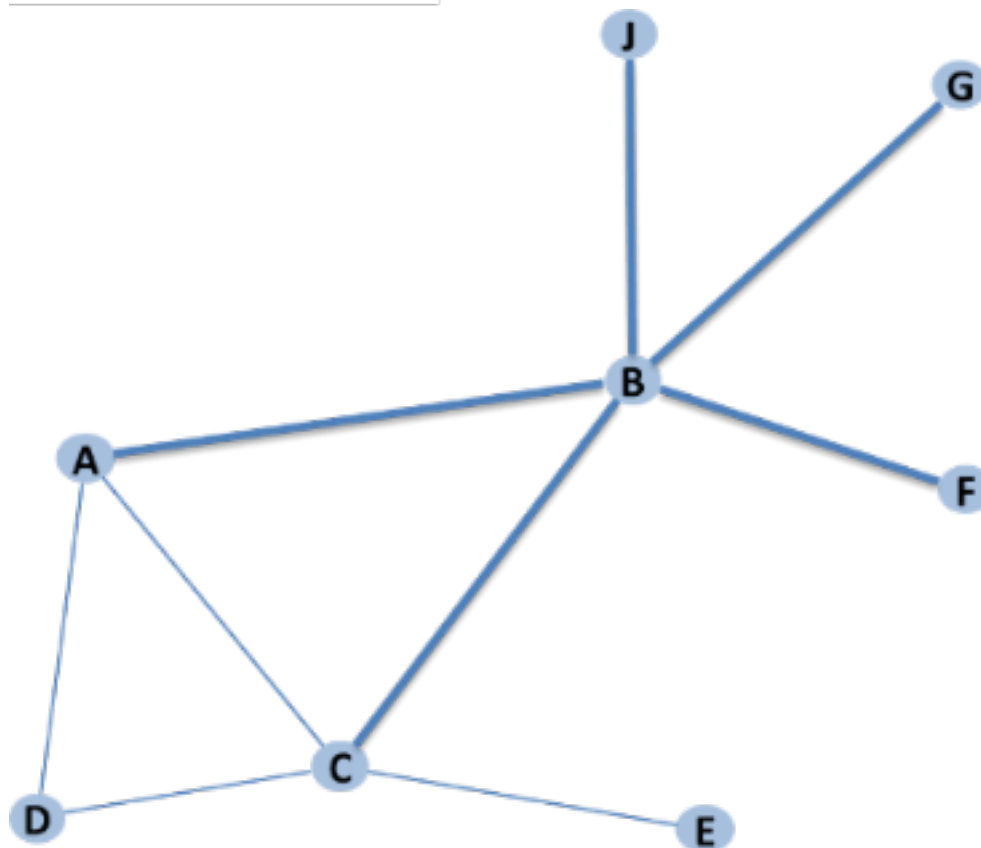
2. Wat is sociaal netwerk analyse?

Een sociale netwerk analyse brengt relaties en stromen tussen mensen, groepen, organisaties en andere informatiebronnen die met elkaar verbonden zijn in kaart. Door middel van deze analyse kunnen relaties en onderlinge verhoudingen in kaart worden gebracht tussen zogenoemde 'knooppunten'. De knooppunten in deze context zijn de bestemmingen, dit kunnen attracties, hotels en restaurants zijn. De verbindingen tussen de knooppunten geven de interacties weer. In dit geval zijn deze verbindingen het aantal keren dat beide bestemmingen zijn gereviewd door eenzelfde reviewer. Bijvoorbeeld wanneer een toerist *Het Fries Museum* in Leeuwarden bezoekt en reviewed op Tripadvisor en vervolgens een hapje gaat eten bij restaurant *De Walrus* in Leeuwarden, dan kan er een verbinding worden gemaakt tussen deze activiteit en dit restaurant. In dit geval is er een verbinding tussen de knooppunten *Het Fries Museum* en Restaurant *De Walrus*.

Eén van de belangrijkste toepassingen van sociale netwerk analyse, is de identificatie van belangrijke knooppunten in het netwerk (Wasserman & Faust, 1994). Voor een toeristische bestemming kan belangrijk zijn om te weten wat belangrijke knooppunten zijn. Er zijn drie verschillende rollen gebaseerd op de netwerk structuur: *hubs*, *bruggen* en *omroepers*. De *hub* brengt knooppunten in verbinding, de *brug* is de schakel tussen diverse netwerken en de *omroeper* is de kortste schakel tussen verschillende netwerken. Deze rollen kunnen worden gevonden door middel van centraliteitsmetingen. Voor de mate van 'centraliteit' wordt een onderscheid gemaakt tussen een drietal maatstaven: *degree centraliteit*, *closeness centraliteit* en *betweenness centraliteit*.

1. De hubs van het netwerk: degree centraliteit

Figuur 3: Degree Centraliteit, waarbij B = 5



Degree centraliteit maakt inzichtelijk hoeveel verbindingen een knooppunt heeft met andere knooppunten. Hoe meer directe verbindingen een knooppunt heeft, hoe hoger de *degree centraliteit* van het knooppunt. *Degree centraliteit* brengt het belang van een accommodatie in kaart binnen een netwerk (Liu, 2012), waarbij onderscheid kan worden gemaakt in: in-degree en out-degree centraliteit. *In-degree centraliteit* weerspiegelt het vermogen van een accommodatie om toerisme stromen te verzamelen. Daartegen weerspiegelt *out-degree centraliteit* van een accommodatie de mate waarin de accommodatie in staat is om als toegangsweg te fungeren naar andere accommodaties.

2. De omroeper van het netwerk: closeness centraliteit

Met *closeness centraliteit* wordt gemeten hoe goed een knooppunt is verbonden met andere knooppunten. In feite wordt er gemeten hoe ver de kortste paden van een knooppunt af liggen van alle andere knooppunten. Waarbij de definitie van het kortste pad: de kortste verbinding tussen twee knooppunten is. Hoe hoger de *closeness centraliteit*, hoe dichter het knooppunt bij andere knooppunten staat. Een hoge *Closeness Centraliteit* van een accommodatie, betekent dat de overdraagbaarheid naar andere knooppunten hoog is. Daarnaast is de mogelijkheid groot bij hoge *closeness centraliteit* om de kern van een netwerk te zijn en dus minder snel gedomineerd te worden door andere accommodaties (Liu, 2012).

3. De bruggen van het netwerk: betweenness centraliteit

Betweenness Centraliteit wordt gedefinieerd als het aantal keren dat een knooppunt als een 'brug' dient langs de kortste paden tussen twee knooppunten. Als node A door node B naar node F gaat, dan loopt het kortste pad van node A naar F door B als brug, waardoor de *betweenness centraliteit* score met 1 wordt verhoogd voor node B. Het knooppunt met de hoogste *betweenness centraliteit* heeft de meeste aantal korte paden door zich heen lopen. Deze meting wordt gebruikt om voor een accommodatie de afhankelijkheid van andere accommodaties te meten. Een hogere *betweenness centraliteit* score betekent meer controle over toerisme stromen en meer structurele voordelen, zoals afhankelijkheid van andere accommodaties (Liu, 2012). Om interpretatie te vereenvoudigen wordt er gebruik gemaakt van een *genormaliseerde betweenness centraliteit*, waarbij de scores resulteren tussen 0 en 1.

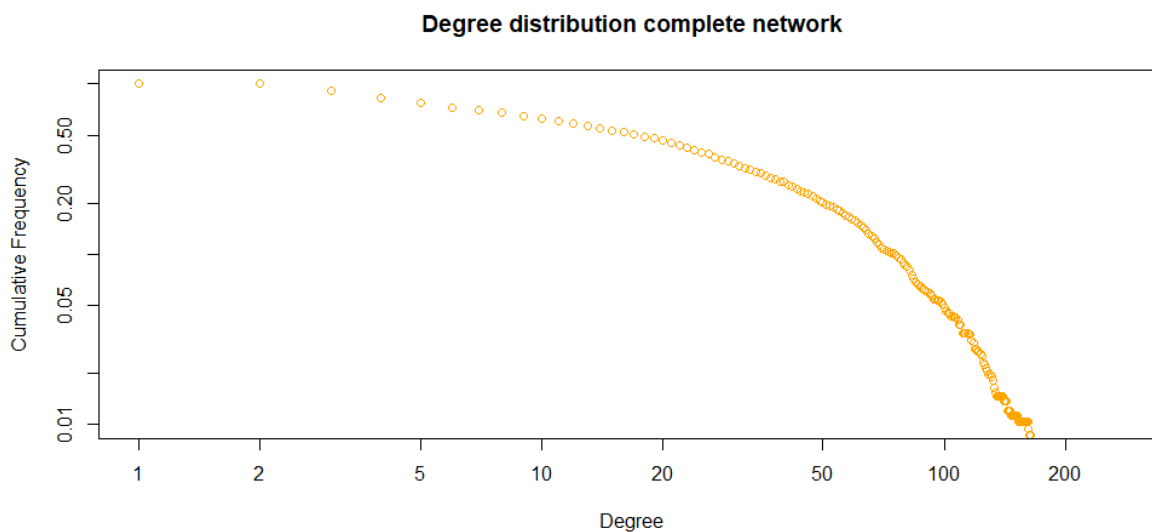
2. Resultaten algeheel netwerk

Netwerkteoretici hebben een arsenaal aan ‘universele’ eigenschappen geïdentificeerd en diverse technieken ontwikkeld waarmee elk abstract netwerk geanalyseerd kan worden. Dit geeft de mogelijkheid om netwerken met elkaar te vergelijken op basis van hun netwerkeigenschappen. Hieronder lichten we toe welke metingen voor deze analyse van belang zijn.

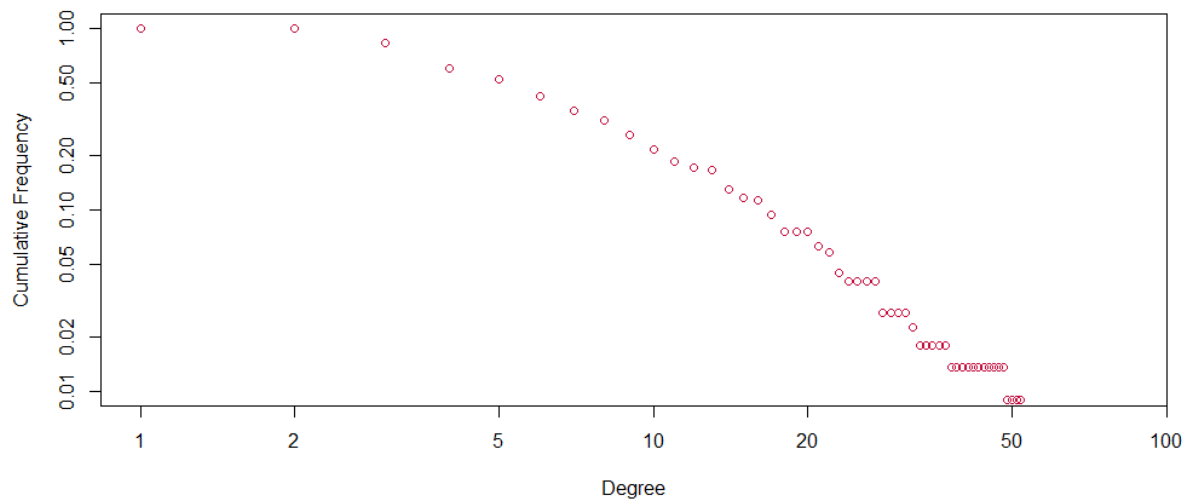
1. Distributiegraad

De distributiegraad is een eerste inzicht wat uit het netwerk kan worden afgeleid. De distributiegraad geeft een beeld van de cumulatieve verdeling van de centraliteit van de knopen in een netwerk. Oftewel de afbeeldingen van figuur 4 geven de verdeling van het aantal verbindingen van een accommodatie in het netwerk weer. In alle vier netwerken is een scheve verdeling naar rechts zichtbaar, waarbij een klein aantal knooppunten veel verbindingen heeft met andere knooppunten, maar het merendeel van de knopen weinig verbindingen heeft. Deze verdeling duidt op de zogenoemde ‘Power law distribution’. (Toubia & Stephen, 2010) Dit is een machtsverdeling waarbij knopen voorkeur hebben om verbinding te maken met knopen van een hoge graad. Het netwerk bestaat uit een voorkeursverbinding, waarbij accommodaties die vaak gereviewed worden – en dus een centrale plek hebben binnen het netwerk - waarschijnlijk vaker nieuwe bezoeken en reviews aantrekken.

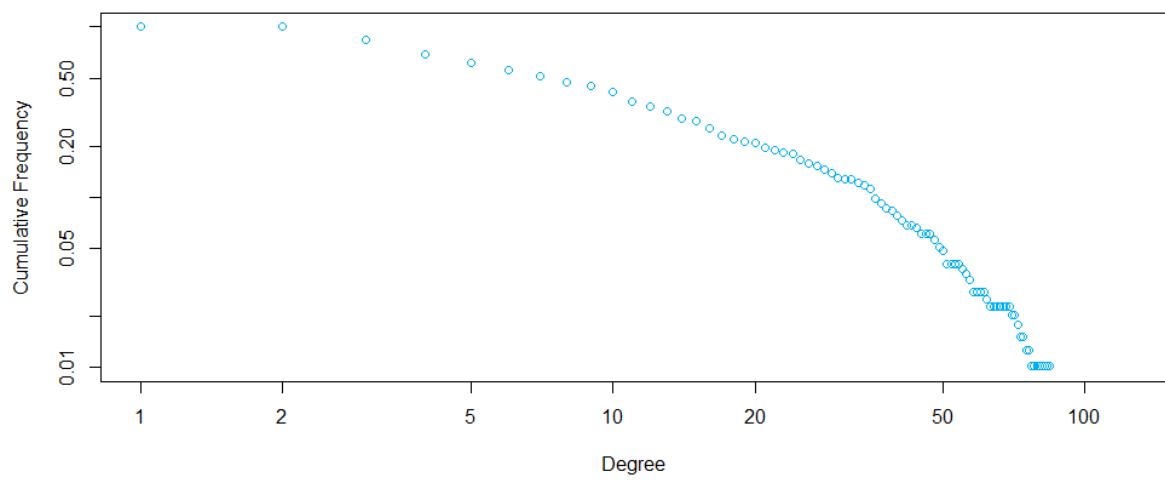
Figuur 4: Distributiegraad van de vier verschillende netwerken

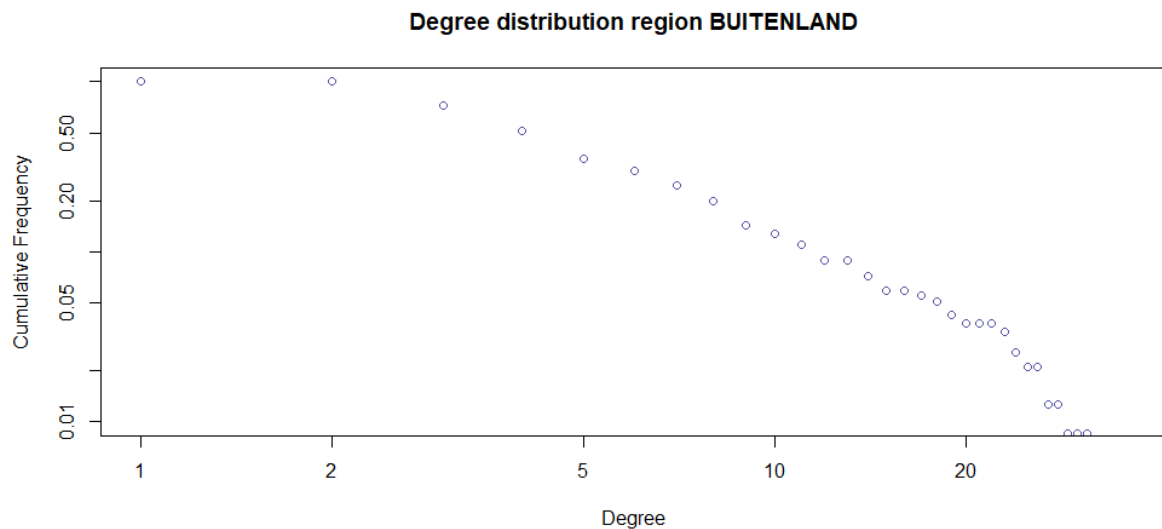


Degree distribution region FRIESLAND



Degree distribution region NEDERLAND





2. Netwerkdichtheid

Elk netwerk bestaat uit knopen en verbindingen. Daarmee is het mogelijk om de netwerkdichtheid te van een netwerk te berekenen. De netwerkdichtheid beschrijft het aantal knopen dat is verbonden in verhouding tot de potentiële verbindingen van de knooppunten. Dit levert een overzicht aan beschrijvende statistieken op. De beschrijvende statistieken van de vier onderzochte netwerken worden weergegeven in tabel 1.

[Vergelijking]=[Vergelijking]

Van de vier netwerken, heeft Buitenland (0.004) de laagste netwerk dichtheid. Het complete netwerk heeft de hoogste netwerkdichtheid (0.012). Dit geeft aan dat het bezoekerspatroon van het gehele netwerk binnen Friesland vergelijkbaar is. De dichtheid tussen 0.1 en 0.01 wordt gezien als een hoge dichtheidsgraad van het netwerk. Dit geeft aan dat het reviewgedrag op Tripadvisor van bezoekers in Friesland restaurants, accommodaties en hotels elkaar verbinden in Friesland.

Netwerk	Network Density	Knopen	Verbindingen	Transitiviteit
Complete Network	0.012280	1160	16510	0.1843897
Friesland	0.006943549	845	4952	0.1246944
Nederland	0.009112803	1005	9195	0.1823919
Buitenland	0.004225301	708	2115	0.1413434

3. Transitiviteit

De transitiviteit van een netwerk is een andere vorm van het meten van de dichtheid van een netwerk. Echter, deze vorm van dichtheid kijkt gericht naar de vraag: in hoeverre mogen we verwachten dat vrienden van vrienden zelf vrienden zijn. Oftewel, wat is de kans dat wanneer er een verband is tussen a-b en b-c, dat a-c ook verbonden zijn.

[Vergelijking]

Een normale waarde voor transitiviteit is tussen 0.3 en 0.6. Voor alle vier de netwerken van accommodaties in Friesland is de transitiviteit laag. Dit betekent dat veel accommodaties niet gereviewed – en wellicht dus ook niet bezocht – worden die wel binnen eenzelfde cluster van gelijke activiteiten kunnen zitten. De mate van transitiviteit heeft invloed op de wijze van beïnvloeding van de accommodatie bezoeken. In dit geval betekent

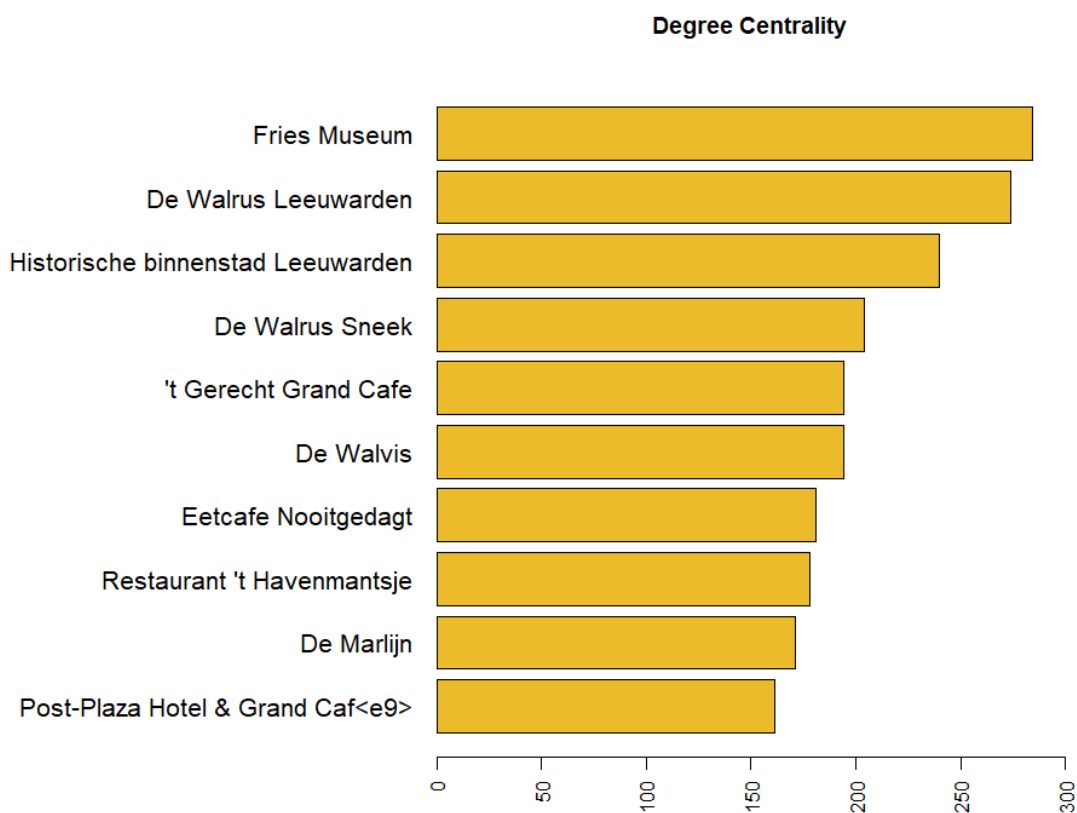
het dat je direct contact moeten zoeken tussen knooppunten en dat toeristen en bezoekers niet op de hoogte zijn van alle activiteiten en accommodaties binnen een cluster.

3. Resultaten knooppunten over algehele netwerk

1. Degree Centraliteit - De HUBS van het netwerk

De centraliteit graad in een netwerk is gelijk aan het aantal accommodaties met wie deze accommodatie in het netwerk een relatie onderhoudt. Met andere woorden, de graad is gelijk aan het aantal directe relaties van de accommodatie in het netwerk, en wordt vaak aangeduid als 'hub'.

De gemiddelde degree centraliteit in Friesland is 16, dit houdt in dat elke accommodatie die is beoordeeld een connectie heeft met 16 andere accommodaties. Zoals in figuur 5 is te zien, heeft het Fries Museum het meest aantal verbindingen (284). Daarnaast is te zien dat de top drie centrale knooppunten in het gehele netwerk, zich alle drie in Leeuwarden bevinden.



Figuur 5: Degree Centraliteit gehele netwerk

Sum	33226
Mean	16
Variance	1207,91

Standard Deviation	34,75
Min.	1
Max.	284

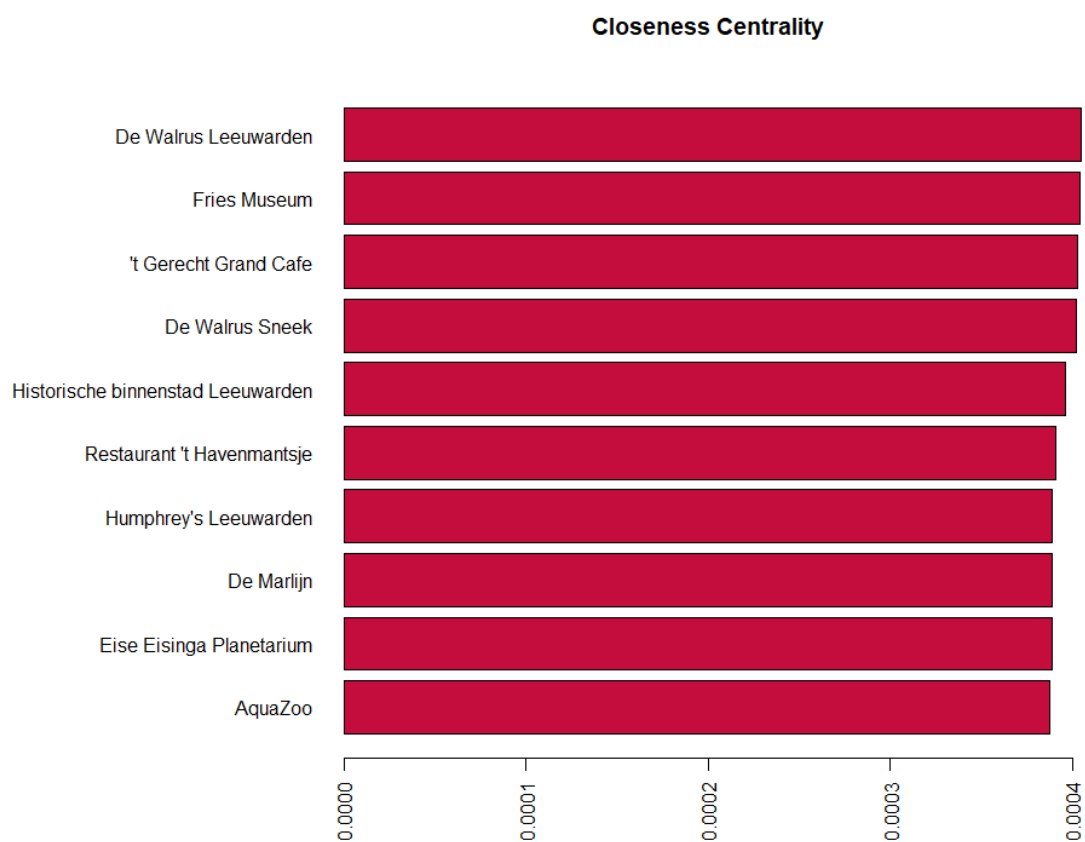
Figuur 6: Descriptives Degree Centraliteit

2. Closeness Centraliteit

In figuur 7 worden de tien accommodaties met de hoogste closeness centraliteit, oftewel nabijheid, weergegeven. De waarden van de closeness centraliteit liggen allemaal dicht bij elkaar en zijn erg laag. De vergelijkbare closeness centraliteit scores tussen de knooppunten duidt op een netwerk met een hoog aantal verbindingen. In de huidige context wordt er gekeken naar 1250 accommodaties, waardoor closeness centraliteit minder nuttig is. Om daadwerkelijk invloedrijke gebieden op te sporen, is deze meting waardevoller binnen kleinere clusters.

In het complete netwerk hebben De Walrus in Leeuwarden, het Fries Museum, 't Gerecht Grand Café en de Walrus in Sneek de hoogste closeness centraliteit. Binnen de huidige context hebben deze accommodaties de kortste verbindingen met andere knooppunten en dus een hoge overdraagbaarheid binnen het netwerk. Het zijn de meest toegankelijke accommodaties op basis van de sociale netwerkanalyse.

Figuur 7: Closeness Centraliteit gehele netwerk



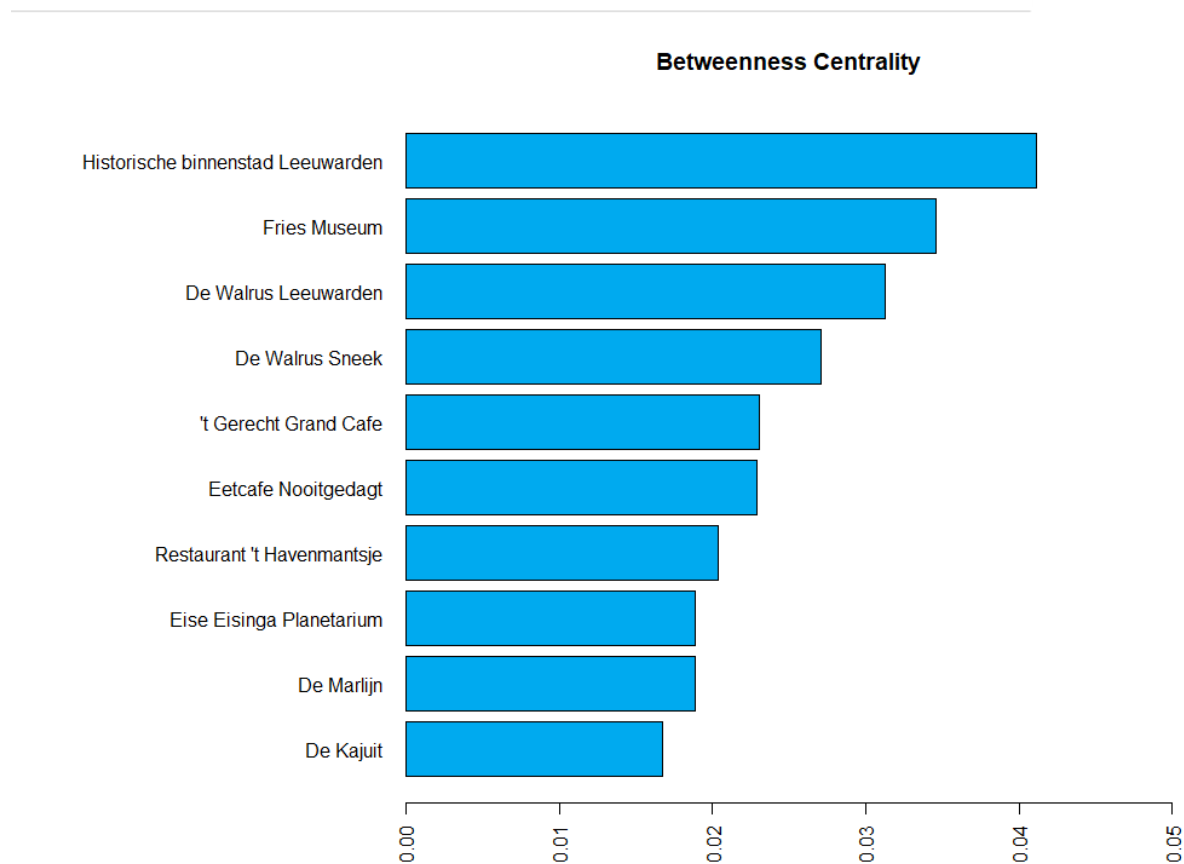
Sum	0.35330920263094
Mean	0. 00030483969166
Variance	0. 00000000165489
Standard Deviation	0. 00004068033376
Min.	0. 00014797277301
Max.	0.0004051863857374

Figuur 8: Descriptives Closeness Centraliteit

3. Betweenness Centraliteit

In figuur 9 worden de tien accommodaties met de hoogste betweenness centraliteit weergegeven. In het complete netwerk heeft de Historische binnenstad Leeuwarden, Het Fries Museum en de Walrus in Leeuwarden deze 'brugfunctie'. Binnen deze context betekent dit dat deze accommodaties het meest voorkomen als 'kortste pad' tussen twee andere knooppunten. Dit betekent dat deze accommodaties meer controle hebben over de toerisme stromen binnen Friesland en daarbij als soort sleutelbruggen binnen het netwerk van accommodaties kunnen fungeren.

Figuur 9: Betweenness Centraliteit algehele netwerk



Sum	1214414
Mean	1048
Variance	5045311
Standard Deviation	2246
Min.	0.0
Max.	27560

Figuur 10: Descriptives Betweenness Centraliteit

4. Visualisaties Netwerken

Dan de netwerken zelf. Door middel van de visualisaties wordt inzichtelijk wat de structuur van het netwerk is en wat de onderlinge verbindingen tussen de knooppunten zijn. Hieronder worden de resultaten weergegeven van de 4 onderscheiden netwerken op basis van de herkomst: het complete netwerk van bezoekers, het netwerk voor Friezen zelf, voor Nederlandse bezoekers en voor de buitenlandse bezoekers.

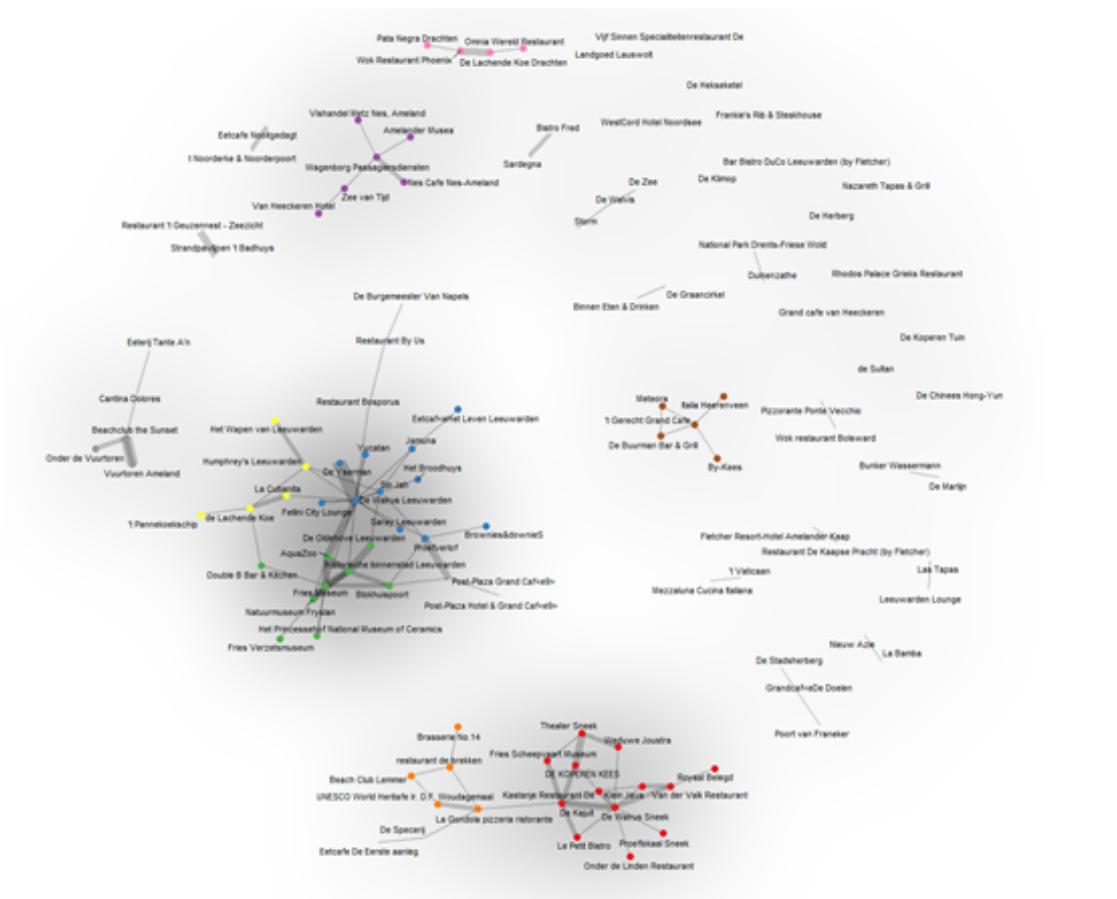
1. Complete Netwerk

Het complete netwerk met bezoekers uit alle landen is gevisualiseerd in afbeelding 11. Om het netwerk overzichtelijker te maken, is er gefilterd op een minimale graad van 10 (oftewel minimaal 10 connecties). Opvallend in dit netwerk is dat het netwerk een aantal subnetwerken heeft wat voornamelijk geografisch is georiënteerd. De gebieden die het meest naar voren komen zijn de eilanden en de stad Leeuwarden.



2. Netwerk herkomst Friesland

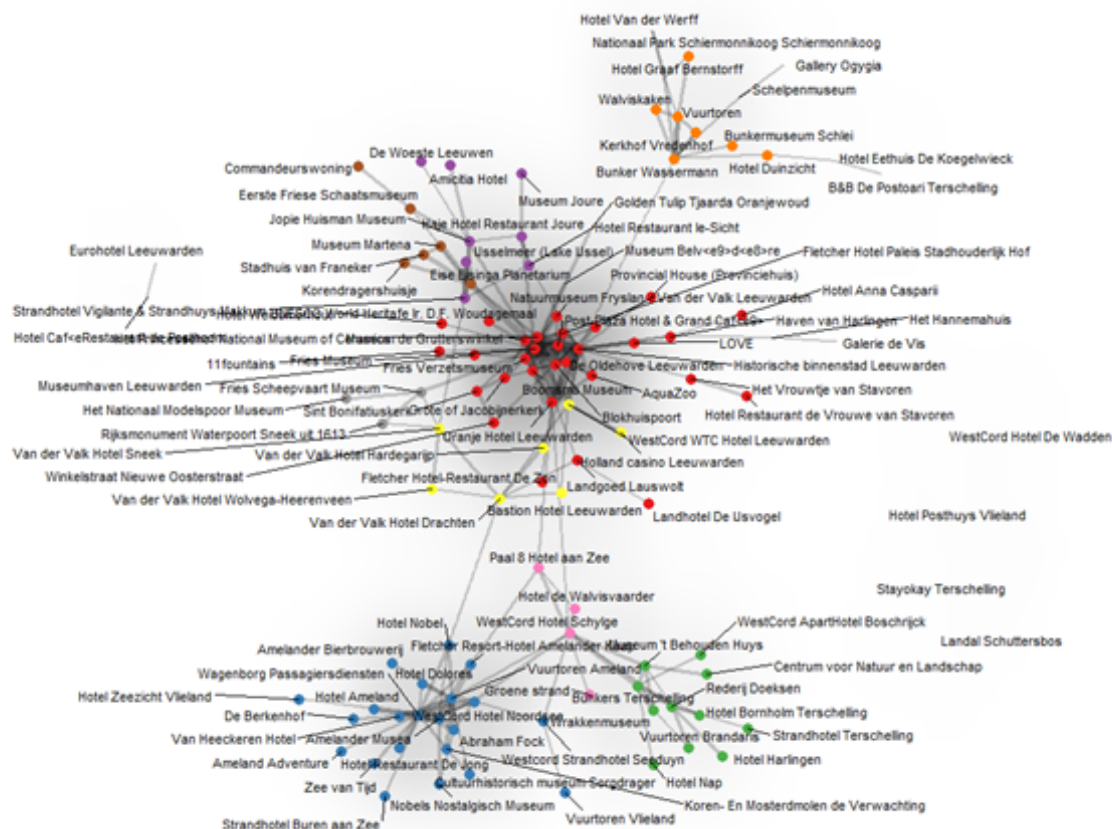
Het netwerk met bezoekers uit Friesland is gevisualiseerd in afbeelding 12. Om het netwerk overzichtelijker te maken, is er gefilterd op een minimale graad van 2. Opvallend in dit netwerk is dat het netwerk een aantal subnetwerken heeft wat voornamelijk geografisch is georiënteerd. Het grootste subnetwerk bestaat uit accommodaties uit Leeuwarden, daarnaast komen Lemmer en Sneek veel voor.



Figuur 12: Netwerk Herkomst Friesland

3. Netwerk herkomst Nederland

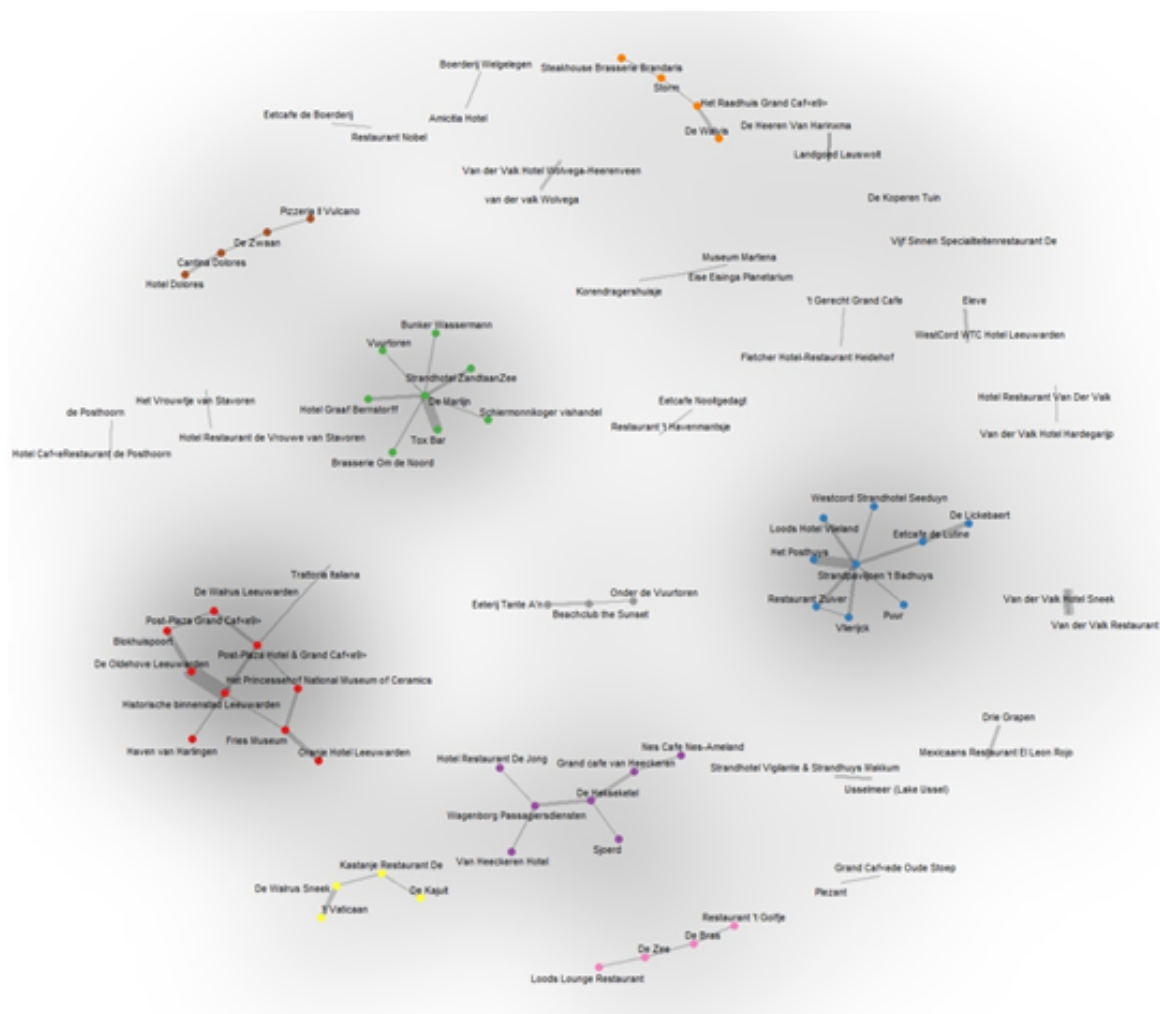
Figuur 13: Netwerk Herkomst Nederland



Het netwerk met bezoekers uit Nederland is gevisualiseerd in afbeelding 13. Om het netwerk overzichtelijker te maken, is er gefilterd op een minimale graad van 2. Vergeleken met de voorgaande twee netwerken is het opvallend dat dit netwerk niet regionaal geclusterd is, maar meer in zijn geheel verbonden is, met de verschillende clusters. De verschillende kleuren geven de verschillende clusters binnen het netwerk weer op basis van community detectie. Deze clusters zijn wederom geografisch georganiseerd.

4. Netwerk herkomst Buitenland

Het netwerk met bezoekers uit het buitenland is gevisualiseerd in afbeelding 14. Om het netwerk overzichtelijker te maken, is er gefilterd op een minimale graad van 2. Opvallend in dit netwerk is dat het netwerk heel veel kleine subnetwerken heeft. Deze subnetwerken zijn geografisch georiënteerd. Dit zou kunnen betekenen dat bezoekers uit het buitenland vaak één plek in Friesland bezoeken en niet rondreizen langs verschillende plekken.



Figuur 14: Netwerk Herkomst Buitenland

5. Bronnen

- Hernández, J., Kirilenko, A., & Steochenkova, S. (2018). Network approach to tourist segmentation via user generated content. *Ann Tour Res*, 35-47.
- Liu, F. &. (2012). Roles and functions of tourism destinations in tourism region of south anhui: A tourist flow network perspective. *Chinese Geographical Science*, 755-764.
- Miguéns, J., Baggio, R., & Costa, C. (2008). Social media and Tourism Destinations: TripAdvisor Case Study. *Advances in Tourism Research*, 26-28.
- Stepchenkova, S., Kirilenko, A., & Hernández, J. (2018). Network approach to tourist segmentation via user generated. *Annals of Tourism Research*, 35-47.
- Toubia, O., & Stephen, A. (2010). Deriving Value from Social Commerce Networks. *Journal of Marketing Research*, Vol 47, Issue 2.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and application*. Cambridge: Cambridge University Press.