

De Programmatuur-Sectie

Graphtheoretische analysemodellen

A.J.A.Felling en Th.M.C.M. van der Weegen.

Binnen de 'community power' theorie¹ werd in het begin van de zeventiger jaren de nadruk gelegd op het ontwikkelen van een sociaal-netwerktheorie. Lokale machtsstructuren werden geconceptualiseerd als gerichte, eventueel gewaardeerde, (multi-) graphs. Een operationeel analytisch instrumentarium, middels welke de belangrijkste componenten van een lokale machtsstructuur kunnen worden onderzocht en beschreven, ontbrak evenwel. In het kader van die lokale machtsstudies zijn een aantal graphtheoretische computerprogramma's vervaardigd met behulp waarvan een systematische beschrijving gegeven kan worden van een aantal structurele parameters, zowel op het niveau van een individuele aktor, als op het niveau van de machtskonfiguratie als geheel.

Het onderzoeksinstrumentarium lijkt toepasbaar niet alleen op de problematiek van lokale macht, maar ook op andere probleemgebieden, waar relaties tussen eenheden relevant zijn.

De belangrijkste programma's welke bij het Sociologisch Instituut te Nijmegen momenteel operationeel zijn betreffen SOCMAT, NCLIQUE, HUBBELL, HIMAAT, SEGDI, STATUS, en DYADE. In het onderstaande wordt in het kort aangegeven over welke thematiek de vervaardigde computerprogramma's handelen. Voor een meer uitgebreide behandeling zij verwezen naar de studie van Felling, getiteld 'Sociaal-netwerkanalyse' (Samsom, Alphen aan den Rijn, 1974)².

1. SOCMAT

Dit is een algemeen programma voor gerichte graphs met behulp waarvan een aantal matrixbewerkingen kunnen worden uitgevoerd. Uitgegaan wordt van een adjacentiematrix $A=(a_{ij})$ behorende bij een gerichte graph $G=(X,U)$.

In de adjacentiematrix van de gerichte graph staat aangegeven of er een direkte pijl bestaat in de graph van een punt x_i naar een punt x_j . Uit deze adjacentiematrix kan de uitgraad $od(x_i)$ en de ingraad $id(x_i)$ worden verkregen, verwijzend naar het aantal pijlen waarmee punt x_i incideert en waarvan punt x_i het begin- respectievelijk het eindpunt vormt.

In de bereikbaarheidsmatrix $R=(r_{ij})$ staat aangegeven of er in de graph een pad bestaat van punt x_i naar punt x_j . In SOCMAT wordt daarbij gebruik gemaakt van het theorema: zij A de bij een gerichte graph behorende adjacentiematrix en zij $T=A^k$ dan is t_{ij} het totaal aantal paden in de graph met lengte k . Vanuit de bereikbaarheidsmatrix zijn bepaalbaar zowel de bereikbaarheidsverzamelingen $R(x_i)$: de verzameling van alle punten in de gegeven graph welke vanuit punt x_i kunnen worden bereikt, als de antecedente verzamelingen $Q(x_i)$: de verzameling van alle punten in de gegeven graph van waaruit punt x_i bereikt kan worden.

De bereikbaarheidsmatrix wordt voorts onder meer gebruikt om de sterke en zwakke komponenten in een graph te bepalen, alsmede om de connectednessmatrix C te konstrueren.

Deze connectednessmatrix $C=(c_{ij})$ van de graph bevat de strikte connectednesscategorie tussen punt x_i en punt x_j . Deze connectednesscategorie verwijst naar de graad van samenhang tussen tweetallen punten x_i en x_j , welke aangeeft op welke wijze tweetallen punten met elkaar verbonden respectievelijk voor elkaar bereikbaar zijn. Vanuit de connectednessmatrix kunnen de unilaterale komponenten van de graph bepaald worden. De connectednessmatrix is in het computerprogramma herordend naar de graad van samenhang van de komponenten van sterk, unilateraal naar zwak.

De distantiematrix $D=(d_{ij})$ van de gerichte graph geeft de lengte aan van het kortste pad tussen twee punten x_i en x_j . Uit de distantiematrix kunnen tal van structurele parameters worden bepaald, zoals de diameter van de graph, puntcentraliteiten, alsmede verschillende typen van statusskores. Ook het bepalen van de liaison persons, zijnde artikulatiepunten waarvan de eliminatie uit de graph een 'disconnected' graph oplevert, kan vanuit de informatie bevat in de distantiematrix geschieden. Bij de condensatiematrix $A^M=(a^M_{ij})$ van de gerichte graph korresponderen de rijen en kolommen met de sterke komponenten en korresponderen de cellen met het al dan niet bestaan van een pijl van sterke komponent X_i naar sterke komponent X_j . Vanuit de condensatiematrix kunnen de basispuntenverzamelingen (een minimum aantal punten van waaruit alle punten bereikbaar zijn) en de kontrabasispuntenverzamelingen (een minimum aantal punten

welke vanuit alle punten in de graph bereikt kunnen worden) worden bepaald.

2. NCLIQUE

Via dit programma kunnen alle n -cliques in een gerichte graph bepaald worden. In vele sociaal-wetenschappelijke studies wordt bij het bestuderen van de clique-struktuur volstaan met het bepalen van de zwakke cq. unilaterale cq. sterke komponenten binnen een graph. Deze komponenten geven immers de maximale deelgraphs aan, welke zich kenmerken door een bepaalde eigenschap van verbondenheid en bereikbaarheid. Luce³ heeft het n -clique begrip geïntroduceerd op basis waarvan meer verfijnde clique-detektiemodellen zijn ontwikkeld. Meer stringente eisen worden gesteld aan de bereikbaarheid dan geschiedt bij de konstruktie van een sterke komponent namelijk door eisen te stellen aan de lengte van een spoor tussen elk tweetal punten, die tot eenzelfde clique worden gerekend. Het programma bepaalt alle n -cliques: een zwakke n -clique is een maximale deelgraph van tenminste drie punten met de eigenschap, dat voor elk tweetal verschillende punten x_i en x_j , geldt dat de distantie $d(i,j) \leq n$. Dit wil zeggen dat tot een zwakke n -clique worden gerekend al die punten, die via een pad van tenminste q (waarbij $1 \leq q \leq n$) wederzijds voor elkaar bereikbaar zijn.

Bij een sterke n -clique geldt als extra eis dat elk punt via een pad ter lengte q (waarbij $1 \leq q \leq n$) voor zichzelf bereikbaar is. De verzameling van alle n -cliques van een gerichte graph zijn hiërarchisch te ordenen (met behulp van de inclusierelatie) van de sterke komponenten tot en met de sterke 1-cliques.

3. HUBBELL

In het cliquedetektie-model volgens Hubbell⁴ worden pijlen, welke tussen punten bestaan, geïnterpreteerd als input- respectievelijk outputkanalen voor het transport van informatie. Via de in het netwerk bestaande pijlen zullen er enerzijds paren zijn, welke sterk met elkaar geassocieerd zijn, anderzijds zullen er paren zijn, welke minder met elkaar geassocieerd zijn. Uitgaande van een graph $G=(X,U)$ ontwikkelt Hubbell voor iedere dyade een associatie-index op basis waarvan hij de betrokken dyade kan kenmerken als major- cq. minordyade. Een major-dyade kenmerkt zich door een relatief hoge associatie-index, een minor-dyade door een relatief lage associatie-index. Hubbell

gaat van de veronderstelling uit, dat twee punten, welke tot eenzelfde clique behoren, in het algemeen door meer n -paden in een graph met elkaar zullen zijn verbonden dan twee punten, welke tot een verschillende clique worden gerekend. Hij bepaalt bovengenoemde associatie tussen twee punten dan ook uit een gewogen som van n -paden tussen deze punten. Een deelverzameling C van X vormt, aldus Hubbell, een clique, als voor iedere x_i en x_j bevat in C , waarbij $x_i \neq x_j$, er in de graph $G=(X,U)$ een direkte of indirekte keten van major-dyaden is van x_i naar x_j , en voorts dat C geen strikte deelverzameling is van een andere deelverzameling van X waarvoor hetzelfde geldt (de eis van maximaliteit).

Hubbell's model gaat uit van een gewaardeerde graph, waarvan de waarde $w_{ij} = v(x_i, x_j)$ de sterkte van pijl (x_i, x_j) inzake de grondrelatie van de graph representeert. Bij het uitwerken van het model wordt gebruik gemaakt van de eigenschap dat de adjacenciematrix W convergent is (Hubbell's kolomnormalisatie).

Daar W konvergent is kan Hubbell de totale sterkte tussen puntenparen x_i en x_j bepalen middels de matrix $Y=(I-W)^{-1}$.

De associatieindex m_{ij} , welke de dyadische relatie typeert, definieert Hubbell als $m_{ij} = \min(y_{ij}, y_{ji})$. Vanuit deze associatie-indexen bepaalt de onderzoeker, via een door hem te kiezen postief criterium ϵ , welke van de dyaden als major respectievelijk als minor worden beschreven. Na het bepalen van de major-dyaden worden de cliques bepaald konform de definitie door het successief aan elkaar knopen van de major-dyaden.

Via het programma HUBBELL kunnen de cliques van een multivalente graph, alsmede de statusskores van de onderscheiden punten bepaald worden overeenkomstig het lineaire input-outputmodel van Hubbell. Het statusgedeelte geeft de skores konform de statusmodellen van Hubbell respectievelijk van Katz respectievelijk van Taylor⁵. Bovendien wordt de gemiddelde status per clique berekend.

In het programma zijn een aantal varianten ingebouwd waaruit de onderzoeker kan kiezen. Deze betreffen o.m. de definitie van de clique (de originele Hubbell-clique versus een clique-definitie welke bovendien nog de eis stelt dat de bij de clique behorende graph compleet is d.w.z. dat voor alle tweetallen punten x_i en x_j geldt dat het geordend paar (x_i, x_j) een major-

dyade vormt); verschillende keuzemogelijkheden inzake het criterium E met behulp waarvan de major-dyaden worden bepaald, alsmede keuzemogelijkheden betreffende de lengte van de sequenties welke men in het input-outputmodel wil toepassen. Op welke wijze de Hubbell-cliques hiërarchisch zijn te ordenen en op welke wijze deze samenhangen met de onderscheiden n -cliques staat schematisch weergegeven in Felling's studie Sociaal-netwerkanalyse, p. 138. Het zou te ver voeren in het kader van dit informatieve artikel hier nader op in te gaan.

4. HMAAT

In dit programma zijn een aantal maten opgenomen, die in de literatuur worden voorgesteld om de geleeidheid (hierarchisering) van een configuratie te bepalen, te weten de PC-index van Brams, de IC-index van Russett, alsmede de H-maat en E-maat zoals vermeld in Coleman⁶.

De 'Power Concentration Index' van Brams is gebaseerd op de (non-reciproke) relaties tussen sterke componenten van een graph. Na een hiërarchische niveautoe kenning (waarvoor een aantal procedures zijn aangegeven) van de condensatiegraph G^* , geïnduceerd uit de sterke componenten, te hebben bepaald, kan de onderzoeker via het bepalen van de PC-index danwel van de PC(m,k)-indices de hiërarchisering van de betreffende graph karakteriseren.

De 'Influence Concentration Index' van Russett is onzes inziens als hiërarchiseringsgraad moeilijk te handhaven. Het blijkt een konnektiviteitsindex te zijn, welke de mate aangeeft waarin de transitieve sluiting van de graph afwijkt van een complete graph. De H-maat van Coleman is een 'source-oriented' maat, gebaseerd op de, al dan niet aselekte, verdeling van de uitgaande pijlen over de diverse punten volgens een gelijkmoegelijke multinomiale verdeling. De H-maat is, gezien de berekeningswijze, een functie van de variantie van de (uit)graad.

De E-maat (de klassieke entropie-maat) is in de terminologie van Coleman een 'consequence-oriented' maat, gebaseerd op de hoeveelheid informatie die de verdeling van de (uit)graden van een graph bezit.

5. SEGDIF

Nagegaan wordt in hoeverre het feit dat punten dezelfde cq. verschillende attributen bezitten kovariëert met een bepaalde distributie van de pijlen over de diverse punten.

De distributie van de pijlen over de diverse punten wordt derhalve in verband gebracht met de positie, welke de beide punten in een dyade bezitten ten aanzien van bepaalde attributen. Een tweetal hoofdeffekten van attributen op de verdeling van de pijlen over de diverse punten kunnen worden onderscheiden, te weten:

- een differentiatie-effekt: hiervan is sprake indien meerdere categorieën van punten slechts bereikt kunnen worden door een bepaalde groep van punten;
- een segregatie-effekt: hiervan is sprake indien genoemde attributen de punten in categorieën onderscheidt, waarbinnen wel maar waartussen nauwelijks pijlen bestaan.

In dit computerprogramma wordt het differentiatie- en het segregatie-effekt onderzocht met behulp van de χ^2 -test, terwijl voorts h-maten worden berekend om de mate van het bovenkans respektievelijk onderkans voorkomen van pijlen tussen categorieën van punten aan te geven⁷.

6. STATUS

Uitgaande van een adjacentiematrix van een gerichte graph berekent STATUS voor ieder punt van de graph een statusscore en een nettostatusscore overeenkomstig de modellen van Bavelas, Beauchamp, Sabidussi, Harary, Arney, de conventionele index, MacKenzie, French, Katz, Hubbell en Taylor⁸.

Bij het bepalen van de status van de punten in een graph heeft de onderzoeker via STATUS de keuze uit statusmodellen, welke gebaseerd zijn óf op de distanties, óf op de n-sequenties welke er tussen punten in een graph bestaan, óf uit een combinatie van beide. Het verschil tussen de statusmodellen, gebaseerd op distanties en de statusmodellen, gebaseerd op de n-sequenties, laat zich in een drietal punten samenvatten:

- bij de statusmodellen, gebaseerd op de distanties worden allereerst alle redundante n-sequenties geëlimineerd, zodat slechts de n-paden resteren,

- vervolgens worden bij deze statusmodellen alle niet-kortste paden van punt x_i naar punt x_j geëlimineerd, en

- tenslotte worden alle kortste paden (sporen) van punt x_i naar punt x_j op één na geëlimineerd.

Naast bovengenoemde keuzemogelijkheden, kan de onderzoeker zijn voorkeur uitspreken voor statusmodellen, waarbij slechts rekening

wordt gehouden met uitgaande sequenties cq. paden danwel met statusmodellen, waarin zowel uitgaande als binnenkomende sequenties cq. paden zijn opgenomen. Bij het eerste type resulteert dit in het toekennen van een statusindex per punt; bij het tweede type wordt een nettostatusindex per punt berekend.

Met betrekking tot de bruikbaarheid van deze statusmodellen kan opgemerkt worden dat tot dusverre, althans naar onze mening, het nog nauwelijks mogelijk is een definitieve keuze te maken uit de behandelde analysemodellen, daar relatief weinig onderzoek is verricht naar de validiteit van deze modellen.

7. DYADE

Dit programma betreft schaalanalyse van een multigraph, door middel van een deterministische scalogramanalyse. Uitgegaan wordt van een multigraph $G=(X, U_1, U_2, \dots, U_k)$ waarin staat aangegeven, welke pijlen er tussen de punten bestaan betreffende de onderscheiden grondrelaties $1, 2, \dots, k$. Als analytisch instrument fungeren de, bij de k grondrelaties behorende, gerichte graphs en hun overeenkomstige adjacentiematrices.

Via het DYADE-programma kan een nader inzicht verkregen worden in de patroonmatigheden die zich voordoen in de verdeling van de pijlen over de diverse grondrelaties. Daarbij kan men diverse analyse-eenheden onderscheiden, zoals de afzonderlijke punten en de dyade. Beschouwt men de afzonderlijke punten, dan kan een verder onderscheid gemaakt worden in een subjektschaalanalyse en een objektschaalanalyse.

De subjektschaalanalyse maakt gebruik van de uitgraad in de k adjacentiematrices en bestudeert de patroonmatigheid in de k uitgraden die punten bezitten.

De objektschaalanalyse maakt gebruik van de ingraad in de k adjacentiematrices en bestudeert de patroonmatigheid in de k ingraden die punten bezitten.

De dyadeschaal maakt gebruik van de informatie welke in de elementen van de k adjacentiematrices staat aangegeven. Een dyadeschaal kan al dan niet opgebouwd zijn uit N subjektschalen ten aanzien van één enkel object en/of N personale schalen⁹.

BESLUIT

In het bovenstaande hebben we een kort overzicht gegeven van het bestaande graph-theoretische onderzoeksinstrumentarium,

zoals dat momenteel in de vorm van computerprogramma's operationeel is. Nagenoeg alle analysemodellen en graph-theoretische bewerkingen, welke vermeld staan in de studie 'Sociaal-netwerkanalyse' van Felling (Samsom, Alphen aan den Rijn, 1974) zijn daardoor uitvoerbaar voor graphs met niet meer dan 200 punten. Met uitzondering van NCLIQUE zijn de computerprogramma's helaas nog niet zodanig van aard, dat zij op grote schaal kunnen worden verspreid. Onderzoekers, die gebruik willen maken van genoemde programma's kunnen zich met hun onderzoeksprobleem en bijbehorende data vervoegen bij de Research Technische Afdeling van het Sociologisch Instituut (080-512021).

LITERATUURVERWIJZINGEN.

1. Aiken, M. en Mott, P.E., eds., *The structure of community power*, New York, 1970
 Bachrach, P. en Baratz, M.S., *Power and Poverty. Theory and practice*, New York, 1970
 Bonjean, Ch.M., Clark, T.N. en Lineberry, R.L., eds., *Community politics: a behavioral approach*, New York, 1971
 Clark, T.N., ed., *Community structure and decision-making: comparative analyses*, San Francisco, 1968
 Miller, D.C., *International community power structures; comparative studies of four world cities*, Bloomington 1970
 Felling, A.J.A., *Lokale macht en netwerken. Een methodologische terreinverkenning*, Alphen aan den Rijn, 1974
 Paige, G.D., ed., *Political leadership: readings for an emerging field*, New York, 1972
2. Harary, F., Norman, R.Z. en Cartwright, D., *Structural models: an introduction to the theory of directed graphs*, New York, 1965
 Harris, B., ed., *Graph theory and its applications*, New York, 1970
3. Harary, F. en Ross, I.C., *A procedure for clique detection using the group matrix*, in *Sociometry*, vol.20, no.3, 1957, p.205-215
 Alba, R.D., *A graph theoretic definition of a sociometric clique*, in *Journal of Mathematical Sociology*, vol.3, no.1, 1973, p.113-126
 Luce, R.D. en Perry, A.D., *A method of matrix analysis of group structure*, in *Psychometrika*, vol.14, no.1, 1949, p.95-116
 Luce, R.D., *Connectivity and generalized cliques in sociometric group structure*, in *Psychometrika*, vol.15, no.2, 1950, p.169-190
4. Hubbell, Ch.H., *An input-output approach to clique identification*, in *Sociometry*, vol.28, no.4, 1965, p.377-399
5. Katz, L., *A new status index derived from sociometric analysis*, in *Psychometrika*, vol.18, no.1, 1953, p.39-43
 Taylor, M., *Influence structures*, in *Sociometry*, vol.32, no.4, 1969, p.490-502

6. Brams, S.J., Measuring the concentration of power in political systems, in *American Political Science Review*, vol.62, no.2, 1968, p.461-475
- Brams, S.J., The search for structural order in the international system: some models and preliminary results, in *International Studies Quarterly*, vol.13, no.3, 1969, p.254-280
- Coleman, J.S., Measures of structural characteristics, in *Introduction to mathematical sociology*, New York, 1964, p.430-468
- Russett, B.M., Probabilism and the numbers of units affected: measuring influence concentration, in *American Political Science Review*, vol.62, no.2, p.476-480
7. Coleman, J.S., Relational analysis: the study of social organizations with survey methods, in *Human Organization*, vol.17, no.4, 1958, p.28-36
8. Bavelas, A., A mathematical model for group structures, in *Applied Anthropology*, vol.7, no.3, 1948, p.16-30
- Beauchamp, M.A., An improved index of centrality, in *Behavioral Science*, vol.10, no.2, 1965, p.161-163
- Sabidussi, G., The centrality index of a graph, in *Psychometrika*, vol.31, no.4, 1966, p.581-603
- Harary, F., Status and contrastatus, in *Sociometry*, vol.22, no.1, 1959, p.23-43
- Arney, W.R., A refined status index for sociometric data, in *Sociological methods and research*, vol.1, no.3, 1973, p.330-345
- MacKenzie, K.D., Structural centrality in communications networks, in *Psychometrika*, vol.31, no.1, 1966, p.17-25
- French, J.R.P.Jr., A formal theory of social power, in *Psychological Review*, vol.63, no.3, 1956, p.181-194
- Katz, L., A new status index derived from sociometric analysis, in *Psychometrika*, vol.18, no.1, 1953, p.39-43
- Hubbell, Ch.H., An input-output approach to clique identification, in *Sociometry*, vol.28, no.4, 1965, p.377-399
- Taylor, M., Influence structures, in *Sociometry*, vol.32, no.4, 1969, p.490-502
9. Riley, M.W., Toby, J., Cohn, R. en Riley, J.W.Jr., Scale analysis of collective data, in Riley, M.W., *Sociological Research, I: A case approach*, New York, 1963, p.440-455
- Riley, M.W., Riley, J.W. en Toby, J., *Sociological studies in scale analysis: applications, theory, procedures*, New Brunswick, 1954