UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2017:29

Decentraliserat eller centraliserat passningsnätverk

en studie om relationen mellan passningsnätverk och resultat i fotbollsmatcher

Joakim Rundin

Examensarbete i matematik, 15 hp
Handledare: David Sumpter
Examinator: Jörgen Östensson
Juni 2017

Department of Mathematics
Uppsala University

Abstract:

A study by Thomas Grund on two seasons of Premier League from 2006 to 2008 showed that teams with a more decentralised passing network, connecting players more equally in their passing, are more successful than teams that have centralised passing network. The purpose of this study is to examine three Premier League teams from the season 2015-2016 and investigate if decentralised passing is more successful. The method of this study consists of markov chain process and eigenvector centrality to identify a possible connection between a passing network structure and the result of a game. The result of this study did not show enough statistic evidence to prove the theory from the study of Thomas Grund.

Keywords: egenvektor centralitet, Markov kedjor, passningsnätverk, centraliserat, decentraliserat

Innehållsförteckning

1. Inledning

1.1 Matematik och fotboll

Fotboll har en förmåga att framkalla en drös av känslor, engagemang och historier hos många människor. Sporten kan ses genom en mängd av olika perspektiv och ett av dem är med matematik. Fotboll och matematik har under de senaste åren utvecklat en starkare relation till varandra. En fotbollsmatch ger alltid tillhörande statistik i form av antal passningar, skott, löpdistans, bollinnehav och mycket mer. Men denna väsentliga statistik kan påverkas av en mängd faktorer och för att verkligen förstå den inneboende matematiken i fotboll behöver vi illustrera den för att undersöka inre strukturer, prestationer och taktiker i spelet.

1.2 Markov kedjor

Denna studie kommer använda sig av det matematiska konceptet Markov kedjor för att undersöka olika passningsmönster.

$$\begin{array}{c c} & \begin{array}{ccc} \text{A} & \text{B} & \text{C} \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{array}$$

Figur1.1, Matris A

$$\begin{array}{c c} & \begin{array}{ccc} \text{A} & \text{B} & \text{C} \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{array}$$

Figur 1.2, Matris P

En Markov kedja sägs sakna minne då sannolikheten för nästa händelse inte påverkas av historien. Låt matris A ovan (figur 1.1) illustrera passningar mellan spelare där siffran 1 visar att en passning är möjlig och siffran 0 visar att en passning inte är möjlig. Raderna i matrisen visar vem man kan passa till och kolonnerna visar vem man kan ta emot passningar ifrån. Notera då att person C kan göra passningar till person A och B men kan endast ta emot passningar från person B. Vidare om varje element i matris A divideras med summan av den tillhörande raden fås en matris P (figur 1.2) med sannolikheten för de olika passningsmöjligheterna. Observera då att person B kan passa till person A eller C med sannolikheten 0,5 vardera. Om person C skulle få passningen och därför vara mer benägen att passa tillbaka till person B skulle händelsen påverkas av det förflutna och därför inte vara en Markov kedja. Det som går att notera är att summan av alla värden i respektive rad av P är lika med 1. Alla värden i matris P är större eller lika med 0 samt mindre eller lika med 1. Matriser med dessa två egenskaper kallas för Markovs kedjeprocess (*Markov chain process*).[1]

1.3 Frobenius teori

En Markov kedja har alltid $\lambda=1$ som ett egenvärde och det existerar alltid en egenvektor med egenvärde 1. Den egenvektorn med tillhörande egenvärde 1 kallas då π och har egenskaperna:

1. $\lambda=1$ är det tillhörande egenvärdet
2. $\pi_i \geq 0$ för alla $i \in S$
3. $\sum_{i \in S} \pi_i = 1$
4. $\pi P = \pi$

De första 3 kraven säger endast att π är en fördelning kring Markov kedjor. Det är det sista kravet som visar att π är invariant under avbildningen P. Vektorn $\pi = (\pi_1, \pi_2, \ldots\ldots)$ kallas då stationär fördelning till matrisen P och illustrerar varje utvald spelares värde där den mest centrala spelaren π_j ges av $\pi_j \geq \pi_i : \forall i$.[2]

---

[1] Klein project blog, *How Google works: Markov chains and eigenvalues,* 2015. Tillgänglig via: http://blog.kleinproject.org/?p=280

[2] Klein project blog, *How Google works: Markov chains and eigenvalues,* 2015. Tillgänglig via: http://blog.kleinproject.org/?p=280

1.4 Syfte och frågeställning

Syftet med denna studie är att jämföra centraliserat mot decentraliserat passningsspel för tre Premier League lag med hjälp av Markov kedjor och egenvektor centralitet. Syftet kommer uppnås med hjälp av följande frågeställning:

- Är det mer framgångsrikt att spela ett decentraliserat passningsspel än ett centraliserat passningsspel?
- Är användningen av Markov Kedjor och egenvektor centralitet relevant för en analys av passningsnätverk?

2. Metod

2.1 Egenvektor centralitet

Metoden i denna studie kommer använda sig av egenvektor centralitet. Det finns flera olika metoder för att undersöka centralitet i ett nätverk där den mest använda är grad centralitet (*degree centrality*). I ett nätverk finns det knytpunkter, så kallade noder, som har ett visst antal kopplingar till varandra. Grad centraliteten undersöker då mängden kopplingar som en viss knytpunkt har till andra närliggande punkter. Den knytpunkten med högst antal kopplingar i nätverket har då ett större inflytande i nätverket.[3]

Istället för att undersöka mängden av kopplingar mellan knytpunkter kollar egenvektor centralitet på värdet mellan dessa kopplingar. Egenvektor centralitet ger varje nod i nätverket relativa poäng. Detta poängsystem baseras på att kopplingar till högt kopplade knytpunkter bidrar mer än kopplingar till knytpunkter med en låg kopplingsgrad. [4]

Om en knytpunkt i:s centralitetsvärde betecknas med π_i kan denna effekt tillåtas genom att göra π_i proportionell mot genomsnittet av *i:s* grannars centralitetsvärde:

[3] Hildorsson, Fredrik, *Scalable Solutions for Social Network Analysis*, Examensarbete, 2009, s.13. Tillgänglig via: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-110548

[4] Franke, Ulrik, *Tekniker för analys av data från webben*, Avdelningen för Informations- och aerosystem, Totalförsvarets forskningsinstitut (FOI), Stockholm, 2012

$$\pi_i = \frac{1}{\lambda} \sum_{j=1}^{n} P_{ij}\, \pi_j,$$

där λ är en konstant och P är den tillhörande matrisen till nätverket. Sedan anges vektorn $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \ldots..)$ som innehåller samtliga värden på centraliteten för varje knytpunkt. Då kan ekvationen ovan skrivas som:

$$\lambda \cdot \boldsymbol{\pi} = \mathrm{P} \cdot \boldsymbol{\pi}$$

Således går det att identifiera att $\boldsymbol{\pi}$ är egenvektor till matris P med egenvärde λ. [5]

2.2 Definition egenvärde och egenvektor

Låt P vara en N×N matris. Ett tal $\lambda \in C$ är ett egenvärde till P om det existerar en nollskild vektor $\pi \in C^N$ så att $P\pi = \lambda\pi$. En sådan vektor π kallas egenvektor till matrisen P.

Metoden för att finna egenvärden till matrisen P ges av rötterna till den karaktäristiska ekvationen:

$$\det(\lambda I - P) = 0,$$

där *I* är identitetsmatrisen N×N till P. Sedan ges egenvektorn B med tillhörande egenvärde λ av de nollskilda lösningarna till det linjära homogena systemet av:

$$(\lambda I - P)\pi = 0.$$

2.3 Tillämpning

Med hjälp av Opta data erhållen från *WhoScored.com* och uträkning i Matlab kommer passningsnätverk räknas ut för de 9 mest centrala spelarna i form av en 9*9 matris där varje rad visar gjorda passningar och varje kolonn mottagna passningar från respektive spelare. Denna passningsmatris A nedan är för de 9 spelare med flest mottagna passningar för Tottenham Hotspur i deras seger med 3-0 mot Norwich City säsongen 2015/2016.

---

[5] Hildorsson, Fredrik, *Scalable Solutions for Social Network Analysis*, Examensarbete, 2009, s.14. Tillgänglig via: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-110548

$$A = \begin{bmatrix} 0 & 8 & 8 & 0 & 4 & 2 & 5 & 11 & 10 \\ 8 & 0 & 14 & 1 & 1 & 2 & 3 & 3 & 13 \\ 4 & 16 & 0 & 3 & 3 & 1 & 5 & 7 & 2 \\ 6 & 0 & 7 & 0 & 5 & 6 & 3 & 6 & 1 \\ 6 & 0 & 0 & 2 & 2 & 5 & 4 & 2 & 5 \\ 0 & 1 & 0 & 6 & 5 & 0 & 3 & 3 & 2 \\ 7 & 1 & 0 & 6 & 5 & 3 & 1 & 6 & 3 \\ 6 & 2 & 9 & 15 & 3 & 1 & 6 & 0 & 0 \\ 11 & 9 & 2 & 0 & 10 & 7 & 3 & 3 & 0 \end{bmatrix}$$

Matris A omformas sedan genom att dividera varje element med summan av den tillhörande raden. I rad 1 är summan lika med 48 och varje element i rad 1 divideras med 48 och skrivs i decimalform med 4 decimaler. Notera att element $P_{1,2}$ blir $\frac{8}{48}$ som är ekvivalent med $\frac{1}{6} = 0{,}1667$ avrundat till 4 decimaler. Då bildas en Markov kedja P där varje rad summerar till 1.

$$P = \begin{bmatrix} 0 & 0.1667 & 0.1667 & 0 & 0.0833 & 0.0417 & 0.1042 & 0.2292 & 0.2083 \\ 0.1778 & 0 & 0.3111 & 0.0222 & 0.0222 & 0.0444 & 0.0667 & 0.0667 & 0.2889 \\ 0.0976 & 0.3902 & 0 & 0.0732 & 0.0732 & 0.0244 & 0.1220 & 0.1707 & 0.0488 \\ 0.1765 & 0 & 0.2059 & 0 & 0.1471 & 0.1765 & 0.0882 & 0.1765 & 0.0294 \\ 0.2308 & 0 & 0 & 0.0769 & 0.0769 & 0.1923 & 0.1538 & 0.0769 & 0.1923 \\ 0 & 0.0500 & 0 & 0.3000 & 0.2500 & 0 & 0.1500 & 0.1500 & 0.1000 \\ 0.2188 & 0.0312 & 0 & 0.1875 & 0.1562 & 0.0938 & 0.0312 & 0.1875 & 0.0938 \\ 0.1429 & 0.0476 & 0.2143 & 0.3571 & 0.0714 & 0.0238 & 0.1429 & 0 & 0 \\ 0.2444 & 0.2000 & 0.0444 & 0 & 0.2222 & 0.1556 & 0.0667 & 0.0667 & 0 \end{bmatrix}$$

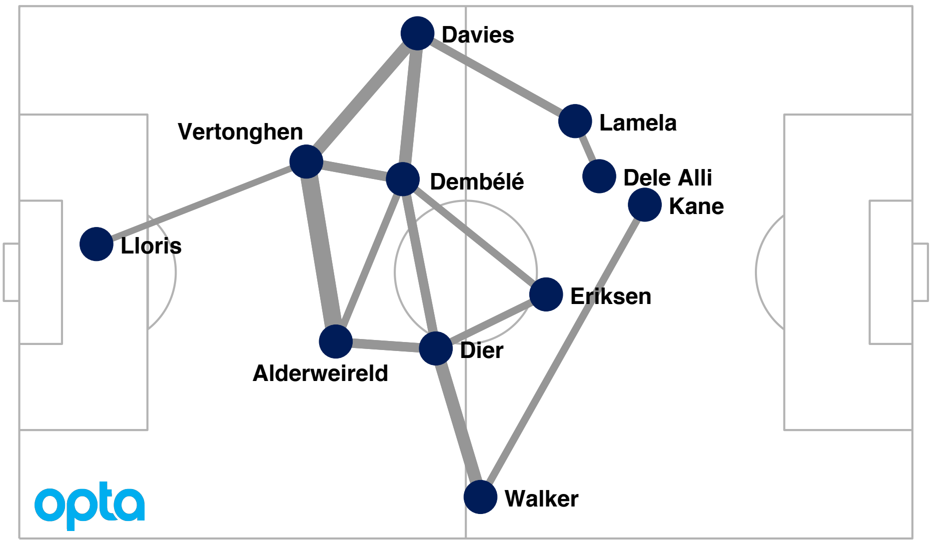
Enligt Frobenius teori har en Markov kedja alltid en egenvektor med tillhörande egenvärde 1. Den tillhörande egenvektorn π för matris P fås genom ekvationen; $(\lambda I - P)\pi = 0$, där egenvärdet $\lambda = 1$ och genom matrisräkning ges den tillhörande egenvektorn:

$$\pi = \begin{bmatrix} 0.4219 \\ 0.3012 \\ 0.3247 \\ 0.3233 \\ 0.3488 \\ 0.2507 \\ 0.3077 \\ 0.3728 \\ 0.3212 \end{bmatrix}$$

För att sedan använda egenvektor centralitet divideras varje element med summan av egenvektorn π för att skapa en stationär fördelning kring matris P. Då fås en ny vektor π där varje element representerar en spelare i passningsnätverket samt $\sum_{i \in S} \pi_i = 1$. Den mest centrala spelaren π_j kan då utläsas enligt $\pi_j \geq \pi_i \ \forall \ i$.

$$\pi = \begin{bmatrix} 0{,}1419 \\ 0{,}1013 \\ 0{,}1093 \\ 0{,}1088 \\ 0{,}1173 \\ 0{,}0843 \\ 0{,}1035 \\ 0{,}1254 \\ 0{,}1081 \end{bmatrix} \quad \begin{matrix} Mousa\ Dembélé \\ Jan\ Vertonghen \\ Toby\ Alderweireld \\ Kyle\ Walker \\ Erik\ Lamela \\ Harry\ Kane \\ Christian\ Eriksen \\ Eric\ Dier \\ Ben\ Davies \end{matrix}$$

Den mest centrala spelaren i denna match för Tottenham Hotspur var Mousa Dembélé som också går att utläsa från en illustration av deras passningsnätverk nedan där Dembélé har en central roll i mitten där en stor mängd passningar gjordes genom Dembélé.



Figur 3.1, Tottenham Hotspur – Norwich City 3-0

2.4 Mest central position

Att det blev Mousa Dembélé som blev den mest centrala spelaren i denna match var ingen ovanlighet. I denna studie med hjälp av egenvektor centralitet är det mittfältare, främst innermittfältare, som har den mest centrala rollen med högst värde i vektorn π. Detta beror främst på att ett fotbollslags passningsspel ofta utgår från innermittfältare som är den positionen som fördelar mest passningar i en match.

2.5 Mått på centralitet

För att sedan undersöka i vilken grad ett lag har haft ett mer centraliserat eller decentraliserat passningsmönster kommer en standardavvikelse räknas ut på samtliga vektorer π för respektive matcher. Standardavvikelsen s_N kommer räknas ut enligt:

$$s_N = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\pi_i - \bar{\pi})^2}. \quad (1)^6$$

där N är antalet element i vektorn π och $\bar{\pi}$ är medelvärdet av samtliga element enligt:

$$\bar{\pi} = \frac{1}{N}\sum_{i=1}^{N}\pi_i. \quad (2)^7$$

Standardavvikelsen av vektorn π för Tottenham Hotspur i matchen mot Norwich City är 0,0161. Om spelarnas värde i vektorn π ligger centrerade kring medelvärdet fås en låg standardavvikelse och det visar på ett mer decentraliserat passningsspel. Om ett passningsnätverk istället genererar en högre standardavvikelse visar det på att en eller några få spelare var mer centrala i passningsnätverket och därmed genererade de ett mer centraliserat passningsspel.

2.6 Mann-Whitney U test

Mann Whitney U- test är ett icke parametrisk test av två oberoende stickprov för att undersöka om det existerar en klar skillnad mellan olika observationspar. Antag att stickprov A består av n_1 observationer av en stokastisk variabel X och stickprov B består av n_2 observationer av en stokastisk variabel Y. Antalet värden fördelat över båda stickproven är då n_1+n_2 stycken. Testet utgår ifrån en nollhypotes H_0 som specificerar på något sätt hur fördelningen mellan observationerna ser ut och ofta antas att variablerna X och Y har samma fördelning som nollhypotes. Testet inleds med att alla observationer i båda stickproven sorteras i storleksordning där det minsta värdet ges nummer 1 och det största värdet ges

[6] Tillgänglig via: http://mathworld.wolfram.com/StandardDeviation.html
[7] Tillgänglig via: http://mathworld.wolfram.com/Mean.html

n_1+n_2. Observationerna i stickprov A har då ersatts med rangnummer $r_1,\ldots,r_{n1}$ och stickprov B med rangnummer $r_1,\ldots,r_{n2}$. Samtliga rangnummer summeras upp till en rangsumma R_1 samt R_2 för respektive stickprov. Sedan söker man den statistiska variabeln U som är det minsta värdet av U_1 och U_2 som räknas ut enligt:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \qquad (3)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \qquad (4)$$ [8]

Nästa steg är då att kontrollera det framtagna U-värdet i nedanstående Mann-Whitney U tabell. [9] Tabellen innehåller kritiska värden relaterade till antalet observationer i respektive stickprov med en signifikansnivå på 5 procent. Om det uträknade U-värdet är mindre eller lika med det tillhörande kritiska värdet förkastas nollhypotesen H_0 och om U-värdet är större än det kritiska värdet avfärdas inte nollhypotesen.

Mann Whitney U-test kommer utföras på standardavvikelserna för vinst respektive förlust. Analysen kommer utgå från noll hypotesen H_0 att standardavvikelserna för respektive vinstmatcher och förlustmatcher har samma fördelning. Det innebär att då att nollhypotesen säger att det inte finns någon statistisk skillnad mellan ett centraliserat och decentraliserat passningsmönster.

[8] Tutorials in Quantitative Methods for Psychology [Elektronisk resurs], Université de Montréal, 2008, vol. 4(1), s. 14-16. Tillgänglig via : http://tqmp.org/RegularArticles/vol04-1/p013/p013.pdf

[9] Tillgänglig via: http://www.snabonline.com/Content/SkillsSupport/MathsAndStatsSupport/M0_14S.pdf

3. Avgränsning

3.1 Antal spelare

Passningsnätverk för ett lag under en fotbollsmatch kommer alltid skapas då vi kan förutsätta att alla spelare i startelvan kommer lyckas göra en passning till en lagkamrat. Nätverken kan påverkas av andra yttre faktor som byten eller skador och därför behövs en avgränsning för vilka spelare som ska ingå i passningsnätverket. Först undersöktes passningsnätverk med samtliga spelare från startelvan och inbytta spelare. Dessa passningsnätverk kunde innehålla mellan 12–14 spelare och dess tillhörande matriser analyserades med egenvektor centralitet. Det problem som uppstod vid en analys av samtliga spelare var att spelare med få eller inga passningar alls, ofta inhoppare, påverkade värdet på varje spelares centralitetsvärde. Därför avgränsades undersökningen först till de 11 spelarna i startelvan och dess passningsmatris. Detta resulterade i någorlunda bättre resultat men det uppmärksammades två problem. Dessa två var att målvakten sällan var inblandad i passningsspelet samt tidiga byten kunde göra att avbytare var mer aktiva i passningsspelet. Med det i åtanke kommer passningsnätverken i denna studie bestå av de 9 spelare som mottagit flest passningar, främst de som startade matchen med vissa undantag som skador och tidiga byten. Detta urvalet har gjorts för att utesluta spelare som blir inbytta i slutskedet av matchen och endast hinner göra ett par passningar eller möjligen inga alls.

3.2 Antal lag

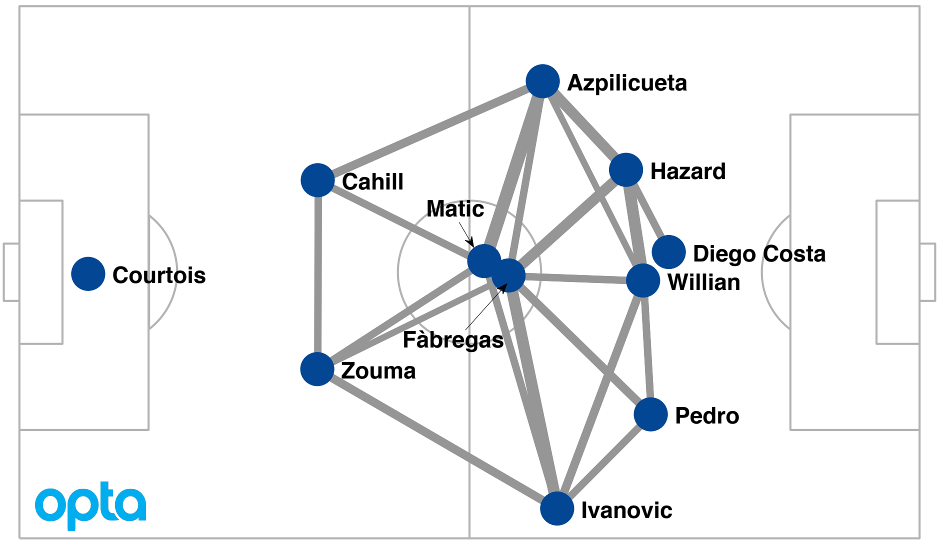
Denna undersökning kommer bestå av de tre Premier League lagen Tottenham Hotspur, Chelsea och Sunderland. Urvalet har gjorts då det kan ge en bredare bild att jämföra lag som slutade i toppen, mitten och botten av tabellen. Tottenham Hotspur var med och kämpade om ligatiteln säsongen 2015/2016 men efter en svag avslutning de sista fyra matcherna slutade de på en tredje plats. Chelsea hade en svag säsong jämfört med säsongen innan och slutade på en 10:e plats. Sunderland har haft det svårt de senaste säsongerna och lyckades precis hålla sig kvar i högsta ligan.

3.3 Antal matcher

För att kunna undersöka om decentraliserat passningsspel är mer framgångsrikt än centraliserat måste en uppdelning göras för att kunna avgöra om en match var framgångsrik eller ej. Därför har denna studie gjort en avgränsning till en jämförelse mellan standardavvikelser för vinst- respektive förlustmatcher. Därmed har de matcher med ett oavgjort resultat filtrerats bort och detta urval har gjorts för att ha möjligheten att visa på en skillnad mellan centraliserat och decentraliserat passningsspel.

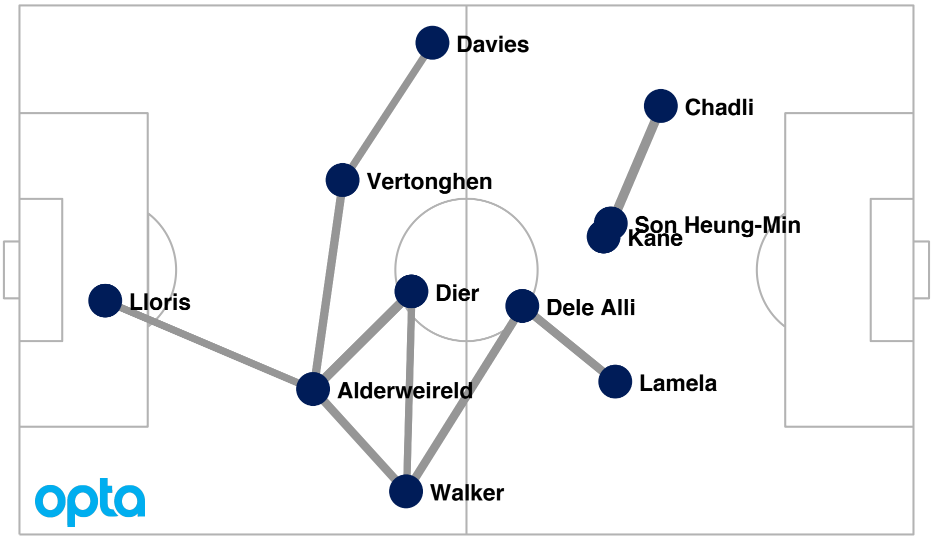
4. Teoretisk bakgrund

En metod att analysera dessa passningsnätverk är genom att identifiera hur centraliserat respektive decentraliserat ett lag spelar. Ett centraliserat nätverk sker när ett lag fokuserar en stor del av sina passningar till ett fåtal aktörer. Figur 3 nedan visar Chelseas passningsnätverk med startelvan hemma mot Crystal Palace säsongen 2015/2016. Varje spelare har en genomsnittlig position där varje mottagen och gjord passning har skett. En linje blir utritad mellan två spelare om minst 10 passningar skett dem emellan. Det som går att notera är att Chelseas nätverk ser ut som ett hjul med två centrala spelare i mitten av hjulet. Fàbregas var inblandad i totalt 131 passningar och Matic i totalt 100 passningar med både mottagna och gjorde passningar i beräkningen. De mesta av Chelseas spel utgår från dessa två spelare då nästan alla andra spelare med 10 eller fler passningar till en lagkamrat är kopplade till Matic eller Fàbregas.



Figur 5.1, Chelsea – Crystal Palace 1-2

Om ett lag istället har ett mer decentraliserat passningsnätverk är samtliga passningar jämnare fördelat över alla spelare. Nedan är Tottenham Hotspurs passningsnätverk mot samma motstånd säsongen 2015/2016. På samma sätt ritas en linje ut mellan två spelare om minst 10 passningar gjorts dem emellan. Det som går att anmärka på Tottenhams passningar är att de är mer jämt fördelade genom laget och hade därför ett mer decentraliserat passningsmönster.



Figur 5.2, Tottenham Hotspur – Crystal Palace 1-0

Enligt en studie utförd av högskolelektorn Thomas Grund 2012 är lag med ett decentraliserat passningsmönster ofta mer framgångsrika än lag som spelar mer centraliserat kring en eller några få spelare. Grund undersökte samtliga 760 stycken Premier League matcher från säsongerna 2006/2007 och 2007/2008 genom att summera samtliga mottagna och gjorda passningar för varje spelare under en match och jämföra de summorna med den spelare i respektive lag som hade flest mottagna och gjorda passningar. [10] Det finns olika metoder att analysera centralitet och flera sådana studier har utförts på passningsnätverk inom fotbollen. Samtliga studier har kommit fram till slutsatsen att ett decentraliserat passningsnätverk gör att spelarna hamnar i ett beroendeförhållande med varandra som kan uppmuntra till ett bättre samarbete och koordination i passningsspelet. Lag som spelar mer decentraliserat blir på samma gång mindre beroende av specifika centrala individer och kan skapa en mer flexibilitet i sitt taktiska spel. [11]

[10] Grund, T.U., Network structure and team performance: The case of English Premier League soccer teams. Soc. Netw. (2012), s, 3. Tillgänglig via: https://doi.org/10.1016/j.socnet.2012.08.004

[11] Grund, T.U., Network structure and team performance: The case of English Premier League soccer teams. Soc. Netw. (2012), s, 3. Tillgänglig via: https://doi.org/10.1016/j.socnet.2012.08.004

5. Analys

I nedanstående tabeller är samtliga standardavvikelser av respektive stationära fördelning π för Tottenham Hotspurs, Chelseas och Sunderlands alla omgångar säsongen 2015/2016. Standardavvikelserna är sorterade i storleksordning från minst till störst i två spalter för respektive lag. De uträknade standardavvikelserna är också färgkodade i tre färger där grönt betyder vinst i matchen, rött innebär förlust och svart text var en oavgjord match.

Tottenham Hotspur säsongen 2015/2016				
Omgång	**Standardavvikelse**		**Omgång**	**Standardavvikelse**
18	0,0161		14	0,0295
10	0,0177		29	0,0306
19	0,0179		35	0,0306
21	0,0183		24	0,0307
20	0,0198		9	0,0309
32	0,0202		37	0,0314
15	0,0229		6	0,0325
11	0,0245		27	0,0326
1	0,0249		31	0,033
23	0,0257		16	0,0331
26	0,0257		2	0,0334
22	0,0258		13	0,0353
36	0,0258		38	0,036
4	0,0259		28	0,0367
7	0,0264		33	0,0373
25	0,0274		34	0,0381
30	0,0278		12	0,0441
3	0,0284		17	0,0497
8	0,029		5	0,0575

Figur 6.1, standardavvikelser för Tottenham Hotspur

Chelsea säsongen 2015/2016				
Omgång	**Standardavvikelse**		**Omgång**	**Standardavvikelse**
16	0,0208		3	0,0444
31	0,0244		15	0,0445
18	0,0266		37	0,0456
5	0,0288		36	0,0461
35	0,0313		29	0,047
10	0,0318		20	0,0478
4	0,0339		12	0,0481
24	0,0386		25	0,0489
23	0,0387		32	0,049
11	0,0391		19	0,0491
34	0,0392		30	0,0493
7	0,0393		2	0,0497
28	0,0395		21	0,0497
6	0,0397		8	0,05
22	0,0415		26	0,0512
13	0,0419		38	0,0522
1	0,0432		9	0,0533
33	0,0437		27	0,0562
17	0,0439		14	0,0591

Figur 6.2, standardavvikelser för Chelsea

Sunderland säsongen 2015/2016				
Omgång	**Standardavvikelse**		**Omgång**	**Standardavvikelse**
38	0,0191		36	0,0429
1	0,0215		35	0,0431
32	0,0231		7	0,0445
25	0,0295		28	0,0447
9	0,0309		12	0,0455
6	0,0314		31	0,0456
26	0,0317		5	0,0459
2	0,0318		4	0,0466
37	0,0339		10	0,0468
33	0,0343		8	0,047
30	0,0354		34	0,047
17	0,0367		20	0,0474
27	0,0368		18	0,0487
16	0,0374		14	0,049
22	0,0379		15	0,0494
13	0,0388		3	0,0498
29	0,0409		21	0,053
23	0,0419		19	0,0555
24	0,0424		11	0,0596

Figur 6.3, standardavvikelser för Sunderland

Vid en första anblick på de tre tabellerna går det inte att urskilja ett tydligt samband där en låg standardavvikelse, med ett mer decentraliserat spel genererar i mer framgångsrika resultat. För att statistiskt säkerställa om det existerar en skillnad i dessa resultat genomförs ett Mann Whitney U-test på två stickprov med vinst- och förlustmatcher. I de tre tabellerna nedan har ett Mann Whitney U-test genomförts på standardavvikelserna för Tottenham Hotspurs, Chelseas och Sunderlands vinst- samt förlustmatcher. Standardavvikelserna är sorterade i storleksordning för vinst- respektive förlustmatcher med antalet observationer n_1 och n_2 angivna. Observationerna har sedan givits rangnummer samt tillhörande rangsummor R_1 och R_2. I varje Mann Whitney U-test gäller det att $n_1 n_2 = U_1 + U_2$ och då måste det ske omräkning i rangordningen om det existerar ekvivalenta observationer som ger likadana rangnummer. Detta måste genomföras för att ovan ekvation ska gälla och det behövdes utföras för Tottenham Hotspurs värden då observation 7 och 8 i storleksordning gav samma rangnummer. Detta löstes genom att ge båda medelvärdet (7+8)/2 = 7,5 som rangnummer.

För Tottenham Hotspur gäller antal vinster $n_1 = 19$ och antal förluster $n_2 = 6$. Samtliga rangnummer finns i tabellen nedan och rangsumman för vinstmatch är $R_1 = 242$ samt förlustmatch $R_2 = 83$. Då fås U_1 och U_2 ut genom:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 19 * 6 + \frac{19(19+1)}{2} - 242 = 62$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 19 * 6 + \frac{6(6+1)}{2} - 83 = 52$$

$$\mathrm{U} = \min(U_1, U_2) = 52$$

Då antas en nollhypotes H_0 som säger att de två stickproven har samma fördelning och därmed går det inte att inte säkerställa en statistisk skillnad mellan decentraliserat och centraliserat. Det uträknade värdet kontrolleras mot det kritiska värdet från ovan Mann Whitney U-tabell med en signifikansnivå 5%. Det tillhörande kritiska värdet är 25 och då 52 > 25 kan nollhypotesen H_0 ej förkastas som gör att det inte går att visa på en skillnad mellan observationerna.

Tottenham Hotspur – Mann Whitney U-test				
Stickprov A Vinster	Stickprov B Förluster	Antal observationer *n*, Rangsumma *R* samt *U*-värde för respektive stickprov	Rangnummer Stickprov A	Rangnummer Stickprov B
0,0161	0,0183	$n_1 = 19$	1	4
0,0177	0,0249	$n_2 = 6$	2	6
0,0179	0,0314	$R_1 = 242$	3	14
0,0245	0,0331	$R_2 = 83$	5	18
0,0257	0,036	$U_1 = 62$	7,5	20
0,0257	0,0367	$U_2 = 52$	7,5	21
0,0258			9	
0,0264			10	

0,0274			11	
0,0278			12	
0,0307			13	
0,0325			15	
0,0326			16	
0,033			17	
0,0353			19	
0,0373			22	
0,0381			23	
0,0497			24	
0,0575			25	

Figur 6.4, Mann Whitney U-test för Tottenham Hotspur

För Chelsea gäller antal vinster $n_1 = 12$ och antal förluster $n_2 = 12$. Dessa värden har givits rangnummer som genererar rangsummorna $R_1 = 171$ och $R_2 = 129$. Genom samma uträkning som ovan fås $U_1 = 51$ och $U_2 = 93$ och då ger det minsta värdet av dessa två U = 51. Det tillhörande kritiska värdet är 37 och då 51 > 37 kan vi ej förkasta H_0 (de två observationerna har samma fördelning) och därmed går det inte visa en statistisk skillnad mellan observationsparen.

Chelsea – Mann Whitney U-test				
Stickprov A Vinster	Stickprov B Förluster	Antal observationer n, Rangsumma R samt U-värde för respektive stickprov	Rangnummer Stickprov A	Rangnummer Stickprov B
0,0313	0,0208	$n_1 = 12$	3	1
0,0387	0,0288	$n_2 = 12$	6	2
0,0395	0,0318	$R_1 = 171$	9	4
0,0397	0,0339	$R_2 = 129$	10	5
0,0419	0,0391	$U_1 = 51$	11	7
0,0439	0,0392	$U_2 = 93$	13	8
0,0444	0,0437		14	12
0,0478	0,0445		17	15
0,049	0,0456		19	16

0,0512	0,0481		22	18
0,0533	0,0497		23	20
0,0562	0,05		24	21

Figur 6.5, Mann Whitney U-test för Chelsea

Sunderland har på samma sätt givits antal vinster $n_1 = 9$ samt antal förluster $n_2 = 17$. Rangsummorna för dessa värden är $R_1 = 134$ och $R_2 = 217$ och genom uträkning ges värdena $U_1 = 64$ och $U_2 = 89$. Därmed fås U = 64 och med det tillhörande kritiska värdet 39 kan vi ej förkasta H_0 eftersom 64 > 39.

Sunderland - Mann Whitney U-test				
Stickprov A Vinster	Stickprov B Förluster	Antal observationer *n*, Rangsumma *R* samt *U*-värde för respektive stickprov	Rangnummer Stickprov A	Rangnummer Stickprov B
0,0317	0,0215	$n_1 = 9$	4	1
0,0339	0,0309	$n_2 = 17$	6	2
0,0354	0,0314	$R_1 = 134$	8	3
0,0388	0,0318	$R_2 = 217$	13	5
0,0468	0,0343	$U_1 = 64$	18	7
0,047	0,0367	$U_2 = 89$	19	9
0,0474	0,0368		20	10
0,049	0,0374		22	11
0,053	0,0379		24	12
	0,0424			14
	0,0445			15
	0,0455			16
	0,0459			17
	0,0487			21
	0,0494			23
	0,0555			25
	0,0596			26

Figur 6.6, Mann Whitney U-test för Sunderland

Det var Mann Whitney U-test för respektive matcher för alla tre lag separat. För att studien ska kunna säkerställa ännu bättre om det existerar en skillnad mellan decentraliserat och centraliserat passningsspel kommer ett Mann Whitney U-test också genomföras på samtliga matcher. Genom att utföra testet på samtliga matcher från alla tre lag kan möjligen Mann Whitney U-testet generera bättre resultat som tyder på en fördel att spela mer decentraliserat.

De undersökta matcherna är tillsammans $n = 70$ stycken matcher med $n_1 = 40$ vinstmatcher och $n_2 = 35$ förlustmatcher. Samtliga matcher gavs en rangordning och det genererade rangsummorna $R_1 = 1487$ och $R_2 = 1363$. Då kan U_1 och U_2 fås genom:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 40 * 35 + \frac{40(40+1)}{2} - 1487 = 733$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 40 * 35 + \frac{35(35+1)}{2} - 1363 = 667$$

Då ger det minsta värdet av U_1 och U_2 att U = 667. I detta fall går det inte kontrollera det kritiska värdet i Mann Whitney U-tabellen, som gjordes i ovan test, då både n_1 och n_2 är större än 20 som är maxgränsen i tabellen. Då kan tillvägagångsättet istället vara att approximera en normalfördelning av U-värdet och göra en uträkning av ett z-värde. Ett z-värde motsvarar en punkt på en normalfördelningskurva och används för att standardisera olika data. Det värdet fås genom:

$$z = \frac{U - m_U}{\sigma_U} \qquad (*)^{12}$$

där m_U är medelvärdet av U_1 och U_2 samt σ_U är standardavvikelsen. Då det är känt att $n_1 n_2 = U_1 + U_2 = 1400$ går det att räkna ut medelvärdet m_U och standardavvikelsen σ_U enligt:

---

[12] The Mann-Whitney U-test -- Analysis of 2-Between-Group Data with a Quantitative Response Variable,2003. Tillgänglig via: http://psych.unl.edu/psycrs/handcomp/hcmann.PDF

$$m_U = \frac{U_1 + U_2}{2} = \frac{n_1 n_2}{2} = \frac{1400}{2} = 700$$

$$\sigma_U = \sqrt{\frac{n_1\, n_2\ (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{1400 * 76}{12}} = 94{,}163$$

Då kan z-värdet beräknas från ekvationen (*) ovan genom:

$$z = \frac{667 - 700}{94{,}163} = -0{,}35$$

För att avsluta Mann Whitney U-testet jämförs det uträknade z-värdet med det kritiska z-värdet, $z_{0,05}$=1,96 för en signifikansnivå på 5% som går att hitta i tabell för kritiska värden.[13] Om absolutbeloppet $|z| = 0.35$ av det erhållna z-värdet är mindre än det kritiska $z_{0,05}$=1,96 kan nollhypotesen H_0 inte förkastas samt om absolutbeloppet $|z|$ är större än det kritiska z-värdet kan vi förkasta H_0 och visa på en skillnad mellan observationerna. I detta fall fås att $|z| < z_{0,05} = 0.35 < 1{,}96$ och nollhypotesen att de två stickproven har samma fördelning går inte att avfärda.

6. Resultat

I denna studie har tre Premier League lag jämförts för att undersöka om det är mer framgångsrikt att spela ett decentraliserat passningsspel än ett mer centraliserat. Analysen genomfördes med hjälp av egenvektor centralitet och Markov kedjor för att beräkna standardavvikelser av samtliga matchers tillhörande vektor π. Resultatet visar genom utförda Mann Whitney U-test att det inte går att statistiskt visa en signifikant skillnad mellan ett decentraliserat och centraliserat passningsnätverk. Samtliga test fick resultatet att nollhypotesen H_0, vinst- och förlustmatcher har samma fördelning, inte gick att avfärda och därför gick det inte att visa en fördel med ett decentraliserat passningsnätverk.

En iakttagelse som går att göra ändå är en jämförelse av alla tre lags medelvärde av standardavvikelser för samtliga matcher, med även oavgjorda resultat inräknade, för att få en grundlig bild hur respektive lag spelat över hela säsongen. Det som går att notera i tabellen

[13] Formelsamling och Tabeller i Statistik och Sannolikhetsteori,2010. Tillgänglig via: http://www.math.kth.se/matstat/gru/5b1508/formler.pdf

nedan är att Tottenham Hotspur hade under säsongen 2015/2016 ett mer fördelat passningsspel bland samtliga spelare än både Chelsea och Sunderland.

Lag	**Medelvärde av samtliga standardavvikelser**
Tottenham Hotspur	0,298
Chelsea	0,428
Sunderland	0,407

6.1 Diskussion

I en fotbollsmatch är det mål och vunna matcher som räknas för att samla in poäng och avsluta i en bra position i tabellen. Om ett fotbollslag använder ett mer decentraliserat passningsnätverk under en match men inte lyckas göra mål på sina avslut och avgöra matcher kommer det leda till att det blir svårt att identifiera ett samband mellan vunna matcher och decentraliserat passningsspel. Detta visualiseras i tabellerna ovan (figur 6.1-6.3) där det inte endast är låga standardavvikelser i vinst matcher. Det finns vinster, förluster och oavgjorda matcher för samtliga lag med både lite lägre och högre standardavvikelser därav blev det svårt att hitta ett samband mellan matchernas standardavvikelse av tillhörande vektor π och resultatet i matchen.

Användningen av Markov kedjor och egenvektor centralitet har hjälpt denna studie att illustrera mönster i passningsspel. Trots att resultatet inte visat på ett statistiskt samband mellan standardavvikelsen och resultatet i matchen är metoden med Markov kedjor och egenvektor centralitet användbar för vidare studier kring fotboll. Ett problem som uppstod med metoden var svårigheten att välja antal spelare i passningsmatrisen. Inledningsvis studerades passningsnätverk med samtliga spelare från startelvan och inhoppare. Spelare med väldigt få eller noll passningar, oftast inhoppare med få spelade minuter, kunde då få höga centralitetsvärden. Sedan undersöktes passningsmatriser med startelvan och det resulterade i något rimligare resultat men både målvakt och tidiga byten kunde påverka värdena i vektorn π. Därför bestod undersökningen av de 9 mest centrala spelarna för att få en bättre bild hur passningsnätverken sett ut under respektive match.

I ett fåtal matcher gick det att belysa en fördel med ett decentraliserat passningsspel som till exempel Tottenham Hotspurs och Chelseas match mot Crystal Palace där Chelsea tydligt hade ett mer centraliserat passningsspel genom två innermittfältare än Tottenham Hotspur som involverade fler spelare i sitt passningsmönster. Tottenham Hotspur vann sin match medan Chelsea förlorade men det är svårt att visa på att deras skillnad i passningsmönster var en orsak till resultaten. För att kunna dra en mer allmän slutsats behöver fler matcher analyseras och jämföras för att kunna visa på en skillnad mellan decentraliserade och centraliserade passningsmönster. I denna studie analyserades 3 lag och 114 matcher för att undersöka kring Thomas Grunds framtagna teori om centralitet. Resultatet visar inte tillräckligt mycket att passningsnätverkens centralitet skulle påverka resultaten i matchen och därav kan ingen allmän slutsats dras att ett decentraliserat passningsnätverk skulle vara mer framgångsrikt. En användning av egenvektor centralitet och Markov Kedjor för en vidare undersökning skulle kunna vara att inkludera samtliga lag i Premier League från säsongen 2015/2016 för att ha möjligheten att jämföra fler matcher och möjligen kunna dra en mer utförlig slutsats kring decentraliserade passningsnätverk och dess fördel.

7. Referenser

Franke, Ulrik, *Tekniker för analys av data från webben*, Avdelningen för Informations- och aerosystem, 2012, Totalförsvarets forskningsinstitut (FOI), Stockholm.

Grund, T.U., Network structure and team performance: The case of English Premier League soccer teams. Soc. Netw, 2012, s, 3. Tillgänglig via: https://doi.org/10.1016/j.socnet.2012.08.004

Hildorsson, Fredrik, *Scalable Solutions for Social Network Analysis*, Examensarbete 30 hp, 2009. Tillgänglig via: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-110548

Klein project blog, *How Google works: Markov chains and eigenvalues,* 2015. Tillgänglig via: http://blog.kleinproject.org/?p=280

The Mann-Whitney U-test -- Analysis of 2-Between-Group Data with a Quantitative Response Variable,2003. Tillgänglig via: http://psych.unl.edu/psycrs/handcomp/hcmann.PDF

Tutorials in Quantitative Methods for Psychology [Elektronisk resurs], Université de Montréal, 2008, vol. 4(1), s. 14-16. Tillgänglig via : http://tqmp.org/RegularArticles/vol04-1/p013/p013.pdf

7.1 Tabeller och Ekvationer

Figur 3.2, Tillgänglig via: http://www.snabonline.com/Content/SkillsSupport/MathsAndStatsSupport/M0_14S.pdf

Ekvation (1), Tillgänglig via: http://mathworld.wolfram.com/StandardDeviation.html

Ekvation (2), Tillgänglig via: http://mathworld.wolfram.com/Mean.html

Formelsamling och Tabeller i Statistik och Sannolikhetsteori, 2010. Tillgänglig via: http://www.math.kth.se/matstat/gru/5b1508/formler.pdf