世界の幸せをカタチにする。
Creating Peace & Happiness for the World
Musashino University
武蔵野大学 学術機関リポジトリ
Musashino University Academic Institutional Repository

Statistical propaties of passing networks in soccer using data from EURO 2016 and LaLiga

メタデータ	言語: jpn 出版者: 公開日: 2022-02-07 キーワード (Ja): キーワード (En): 作成者: 佐藤, 直哉, 木下, 修一 メールアドレス: 所属:
URL	https://mu.repo.nii.ac.jp/records/1667

EURO2016・スペイン一部リーグの サッカーにおけるパスネットワークの統計的特徴

Statistical propaties of passing networks in soccer using data from EURO 2016 and LaLiga

佐 藤 直 哉[1]
Naoya Sato
木 下 修 一[2]
Shu-ichi Kinoshita

1 はじめに

近年，センサー技術の発展やデータベースの整備に伴いスポーツにおいてこれまでとは比較にならない詳細なデータが蓄積され，このようなデータから特定のスポーツ競技に現れる共通の特徴を発見することや，競技者やチームごとの特徴をとらえる研究が盛んになりつつある．このような時間的・空間的に詳細なデータの解析はこれまで自然科学（物理，化学など）の分野で行われてきたため，自然科学で培われた手法が援用される場合も多い．スポーツは大きく個人競技（柔道，陸上競技，水泳，ゴルフなど）と集団競技（サッカー，バスケットボール，ラグビーなど）に分けることができ，個人競技では競技者個人が勝利の目的のために最善の手を尽くすのに対し，集団競技ではチームの勝利に向けて競技者はチームメンバーと協調した行動をとる．多くの集団競技ではゲーム中に監督やキャプテンからの指示に従い競技者個人が行動するだけでなく，競技者個人がピッチ上の味方・敵の位置を把握し最適な行動を自律的に判断し実行している．このように集団競技では構成要素である競技者がお互いに相互作用し，相互作用のありかたが系全体の振る舞いであるチームの勝敗に大きな影響を与えている．このような背景から近年相互作用を表すネットワーク構造と系全体の振る舞いの関係に着目した研究が盛んにおこなわれている．

そのような研究の一つとして，サッカーのパスネットワークに関する研究があげられる．サッカーでは相手にボールを取られないよう，競技者は敵とのバランスをとり自ら動きながら味方の間でボールを回し相手のゴールに迫る．したがって，パスネットワークを考える際に選手のみならず選手の位置も考慮に入れる必要がある．このような観点を踏まえ，現在サッカーのネットワーク研究におけるパスネットワーク作成法は大きく分けて次の 3 つに分類される：(1) 選手をノードとみなした場合，(2) ピッチ上の分割された位置をノードとみなした場

[1] 武蔵野大学大学院工学研究科数理工学専攻修士課程 1 年
[2] 武蔵野大学数理工学センター員 / 武蔵野大学工学部数理工学科教授

合，(3) 選手とピッチ上の分割された位置の組合せをノードとみなした場合の 3 種類である．(1) のネットワークではパスの出し手と受け手の情報により作成されるのに対し，(2), (3) のネットワークではより詳細なピッチ上の選手の位置情報が必要となる．この (3) の視点でサッカーネットワークを解析した研究として T.Narizuka 等 [1] の研究が挙げられる．T.Narizuka 等 [1] は 2014 年ブラジル W 杯アジア予選の 2 試合と EURO2012 の 2 試合と 2011 年キリンチャレンジカップの 1 試合の計 5 試合のパスネットワークの解析を行い，パスの次数と累積度数分布が不完全ガンマ関数で近似することが出来ることを示した．さらに，パスのやり取りをマルコフ連鎖としてモデル化することで，パスの統計的特徴の再現に成功した．この結果は個別の選手や監督の戦術などに依存していないため，サッカーにおいて一般的に成り立つ統計則である可能性がある．ただし，スポーツのような人間を構成要素とする系は自然現象と異なり，構成要素の動きや選手間のコミュニケーション自体が年月とともに変化する可能性があり，2012 年において現れた統計則がその後も成り立ち続けるとは限らない．

そこで本研究では，先行研究で現れた統計的特徴が年月を経ても成り立つ頑健な統計則であるか否かを明らかにするため，同様の手法を用いて EURO2016 およびスペイン一部リーグ (2017/2018) のデータ解析を行い，先行研究との比較を行った．

表 1　本研究で用いるデータで取り扱われている大会・試合数・選手数 [3]

大会名	試合数	イベント数	選手数
スペイン 1 部リーグ	380	628,659	619
イタリア 1 部リーグ	380	647,372	686
ドイツ 1 部リーグ	306	519,407	537
イングランド 1 部リーグ	380	643,150	603
フランス 1 部リーグ	380	632,807	629
2018 年ロシア W 杯	64	101,759	736
EURO2016	51	78,140	552
	1,941	3,251,294	4,299

表 2　本研究の分析で取り扱った EURO2016 の試合（準々決勝・準決勝・決勝）

	試合	日付	結果	ラウンド
(a)	ポルトガル対フランス	2016 年 7 月 10 日	1 - 0 （延長）	決勝
(b)	ポルトガル対ウェールズ	2016 年 7 月 6 日	2 - 0	準決勝
(c)	ドイツ対フランス	2016 年 7 月 7 日	0 - 2	準決勝
(d)	フランス対アイスランド	2016 年 7 月 3 日	5 - 2	準々決勝
(e)	ドイツ対イタリア	2016 年 7 月 2 日	1 - 1 (PK：6 - 5)	準々決勝
(f)	ウェールズ対ベルギー	2016 年 7 月 1 日	3 - 1	準々決勝

2　データの説明

	eventId	subEventName	tags	playerId	positions	matchId	eventName	teamId	matchPeriod	eventSec	subEventId	id
0	8	Simple pass	[{'id': 1801}]	26010	[{'y': 48, 'x': 50}, {'y': 50, 'x': 47}]	1694390	Pass	4418	1H	1.255990	85	88178642
1	8	Simple pass	[{'id': 1801}]	3682	[{'y': 50, 'x': 47}, {'y': 48, 'x': 41}]	1694390	Pass	4418	1H	2.351908	85	88178643
2	8	Simple pass	[{'id': 1801}]	31528	[{'y': 48, 'x': 41}, {'y': 35, 'x': 32}]	1694390	Pass	4418	1H	3.241028	85	88178644
3	8	High pass	[{'id': 1802}]	7855	[{'y': 35, 'x': 32}, {'y': 6, 'x': 89}]	1694390	Pass	4418	1H	6.033681	83	88178645
4	1	Ground defending duel	[{'id': 702}, {'id': 1801}]	25437	[{'y': 6, 'x': 89}, {'y': 0, 'x': 85}]	1694390	Duel	4418	1H	13.143591	12	88178646
...	...	...	...	...	...	...	...	...	...	...	...	...
78135	6		[]	7936	[{'y': 28, 'x': 89}, {'y': 100, 'x': 100}]	1694440	Offside	4418	E2	984.816517		90589035
78136	2	Time lost foul	[{'id': 1702}]	70134	[{'y': 61, 'x': 10}, {'y': 0, 'x': 0}]	1694440	Foul	9905	E2	1026.418188	26	90589357
78137	3	Free Kick	[{'id': 1801}]	70134	[{'y': 63, 'x': 9}, {'y': 78, 'x': 67}]	1694440	Free Kick	9905	E2	1031.553667	31	90589358
78138	1	Air duel	[{'id': 701}, {'id': 1802}]	70410	[{'y': 78, 'x': 67}, {'y': 0, 'x': 0}]	1694440	Duel	9905	E2	1035.342605	10	90589359
78139	1	Air duel	[{'id': 703}, {'id': 1801}]	7936	[{'y': 22, 'x': 33}, {'y': 100, 'x': 100}]	1694440	Duel	4418	E2	1036.319388	10	90589040

78140 rows × 12 columns

図 1　EURO2016 決勝のポルトガルのデータ

この節では本研究で用いたデータの概略について説明する．本研究では，2019 年 10 月にイタリアのガゼッタデロスポーツが主催するスポーツフェスティバルのイベントの 1 つで行われた，サッカーデータチャレンジで一般に公開されたデータを用いる [3]．センサー技術の発展にともない，サッカーのゲーム中における選手の位置やプレーの詳細なデータが取れるようにはなっているが，そのデータは多くの場合公開されない．したがって，上記のようにデータが公開される機会は非常に貴重なものとなっている．今回扱うデータは，2017/2018 シーズンのスペイン，イタリア，ドイツ，イングランド，フランスの 1 部リーグと 2018 年ロシア W 杯と EURO2016 について，試合結果，出場選手，チーム，試合中のイベント，大会，審判，コーチのデータが含まれている．表 1 に各大会の試合数とイベント数を示す. 図 1 に EURO2016 決勝のイベントデータの一部を示す．イベントデータでは，イベント名，イベント Id，サブイベント名，サブイベント Id，試合 Id，チーム Id，選手 Id，イベントが起きた時間，イベントが起きた場所，イベントの成否を表すタグが記載されている．イベント名ではサッカーで起きる大まかなイベントがまとめられており，イベント Id により番号が振り分けられている．サブイベントはイベントをさらに細分化したものである．また，試合 Id では 1,941 試合すべてに番号が振り分けられており，どの試合でイベントが起きたのか記載されている（例えば，EURO2016 決勝は 1694390）．チーム Id と選手 Id も同様に全チーム全選手に番号が振り分けられておりどのチームの誰がイベントを行ったのかがわかる．イベントが起きた時間帯は大きく 2 つに分けられており，1 つ目はさらに前後半，延長前後半，PK 戦の 5 つに分かれている．また，2 つ目はキックオフからの秒数が記載されている．イベントが起きた場所はピッチの左

下を原点にとった座標系を用いて表現され，縦と横の両方とも 0 から 100 で分けられており，イベントが始まった場所の座標と終わった場所の座標が記載されている（縦と横の単位あたりの長さが異なることに注意する）．タグにはそのイベントの成否が記載されている．本データで用いられているイベント Id 等の番号は Wyscout によって定義づけられている．Wyscout とは，イタリアの企業であり，世界中のあらゆるカテゴリーの試合や選手の動画を集めてタグ付けした巨大なデータベースをオンライン上で提供している [4]．今回は先行研究で用いられた EURO2012 のデータと比較するためまず EURO2016 のデータ，続いてスペイン一部リーグのデータ解析を行った．

3　パスネットワークの作成

この節では前節で得られたデータを用いて作成するパスネットワークについて述べる．先行研究と同様の手法を用い，位置と選手の組でノードを決定する．初めにフィールドを 18 分割（ゴール方向に沿って 3 分割，垂直に沿って 6 分割）し，選手がフィールド上のすべての領域を動き回る場合を考え，各エリアに全選手を配置する．つまり，1 チームのノードの最大総数は 198（= 11(選手数)×18(エリア数)）となる．今回のデータでは，イベントが起こった位置を座標で与えているため，この座標とエリアを対応させることで選手とエリアの組で表現されるノードが確定される（図 2 を参照のこと）．続いて，ある選手から別の選手へパスが繋がったとき，対応する（選手とエリアの組で表されている）ノード間を無向エッジで接続する．このとき，エッジのつなぎ方により (a) 同じノード間のパスは何度起こっても 1 つのエッジとみなす場合（パスネットワーク）と，(b) 同じノード間のパスを重複エッジとして取り扱う場合（重複パスネットワーク）の 2 種類のネットワークを作成する．

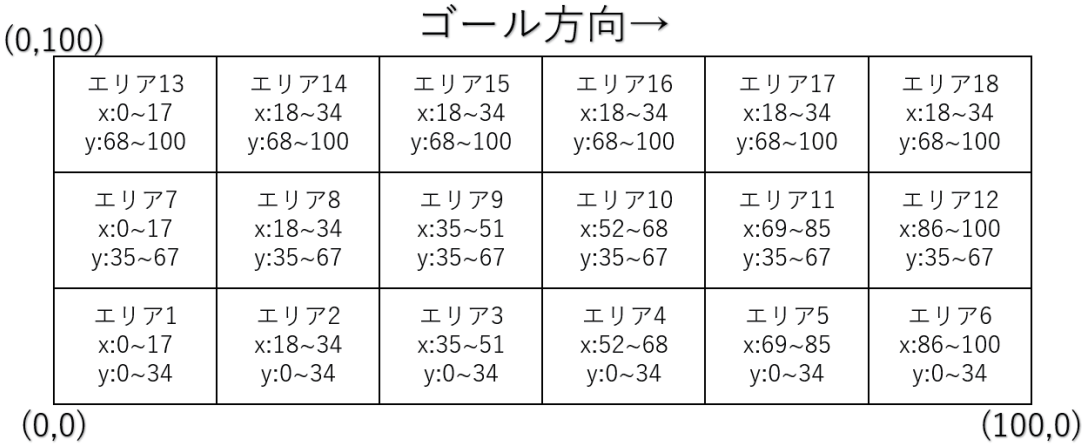


図 2　各エリアの座標範囲

本研究で分析した試合は表 2 の EURO2016 における 6 試合である．また，図 3 はポルトガル対フランスの試合のパスネットワーク例である．今回扱ったデータを用いパスネットワーク

図を作成した場合，非連結グラフになるものあったが，本研究では最大連結部分を用いて解析を行った．このネットワークでは，孤立したノードは除外されており，パスネットワークのノードの約 70 %が平均次数よりも小さい．そのため，平均次数よりとても大きいノードがいくつか存在しており，それらがこのネットワークのハブとしての機能を担っている．パスネットワークの各種統計量は表 3 のとおりである．延長戦や PK で決着がついている試合ではエッジの数が他のものよりも多い．また，平均経路長 l は同サイズのランダムネットワークの平均経路長 l_{rand} とほぼ同等であり，平均クラスター係数 C はランダムネットワークの平均クラスター係数 C_{rand} よりかなり大きい値を示している．これは，サッカーのパスネットワークにはスモール・ワールド性があることを意味しており，先行研究の結果と一致している．

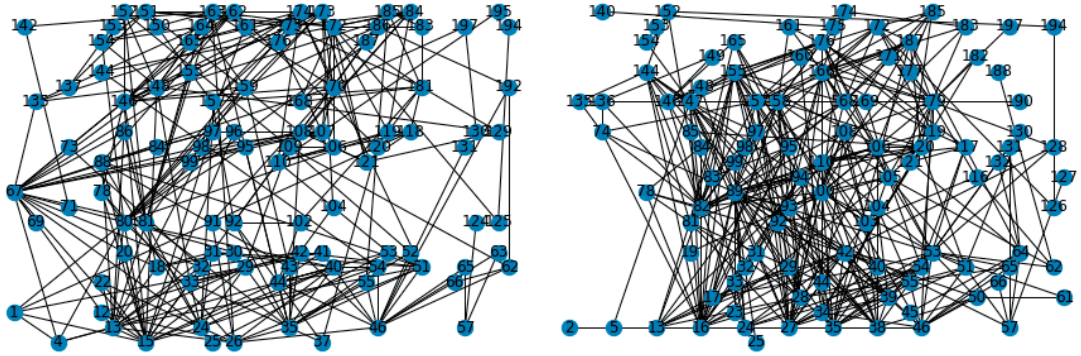


図 3　EURO2016 決勝のネットワーク図．左がポルトガル代表，右がフランス代表 (両図とも右側が攻撃方向である)

4　次数分布

本研究での主要な結果であるパスネットワークの次数の累積度数分布を図 4 に示す．各パネルの曲線はフィッティング関数であり，確率密度関数 $f(k)$ と累積分布関数 $F(k)$ は以下の不完全ガンマ関数で定義される

$$f(k) = \frac{1}{\lambda^{\nu}\gamma(\nu, k_{max}/\lambda, 0)} k^{\nu-1} e^{-\frac{k}{\lambda}}, \tag{1}$$

$$F(k) = \frac{\gamma(\nu, k_{max}/\lambda, k/\lambda)}{\gamma(\nu, k_{max}/\lambda, 0)}. \tag{2}$$

この時，$\gamma(\nu, k_{max}/\lambda, k/\lambda)$ は，k_{max}/λ を上限とする上側不完全ガンマ関数である

$$\gamma(\nu, k_{max}/\lambda, k/\lambda) = \int_{k/\lambda}^{k_{max}/\lambda} t^{\nu-1} e^{-t} dt. \tag{3}$$

k の範囲は $0 \leq k \leq k_{max}$ であり，k_{max}，λ，ν はフィッティングパラメータである．図 4 でのパラメータの値を表 4 に示す．すべての試合において，不完全ガンマ関数はデータをよくフィットしている．ただし，それぞれの試合においてフィッティングパラメータは異なってい

表 3　本研究の分析で取り扱ったパスネットワークの基本統計量 エッジの数 M，平均次数 $\langle k \rangle$，平均経路長 l，平均クラスター係数 C，同一サイズのランダム・ネットワークの平均経路長 l_{rand}，平均クラスター係数 C_{rand}.(l_{rand} が—の部分はランダム・ネットワークの強連結性が保持できなかった)

試合	チーム	M	$\langle k \rangle$	l	C	l_{rand}	C_{rand}
(a)	ポルトガル	523	9.10	3.06	0.18	3.35	0.030
	フランス	668	12.72	2.76	0.24	2.97	0.040
(b)	ポルトガル	358	6.75	3.16	0.16	—	0.017
	ウェールズ	455	6.75	3.00	0.16	3.62	0.018
(c)	ドイツ	592	9.40	3.08	0.24	3.15	0.038
	フランス	241	4.77	3.62	0.17	—	0.0097
(d)	フランス	605	11.10	2.92	0.17	3.13	0.029
	アイスランド	350	6.09	3.47	0.15	—	0.021
(e)	ドイツ	738	12.10	2.98	0.21	2.84	0.038
	イタリア	449	8.55	2.94	0.14	3.62	0.025
(f)	ウェールズ	370	6.79	3.18	0.12	—	0.024
	ベルギー	433	7.27	3.14	0.15	3.73	0.017

る．（1）式より，関数の裾の特性つまり次数 k が大きい部分は k_{max} の値が一番効いてくるが，その他にパラメータ λ が効いてくる．つまり，λ が小さいほど，裾の長い分布になり他の多くの選手・場所にパスを出す選手が存在することを意味している．

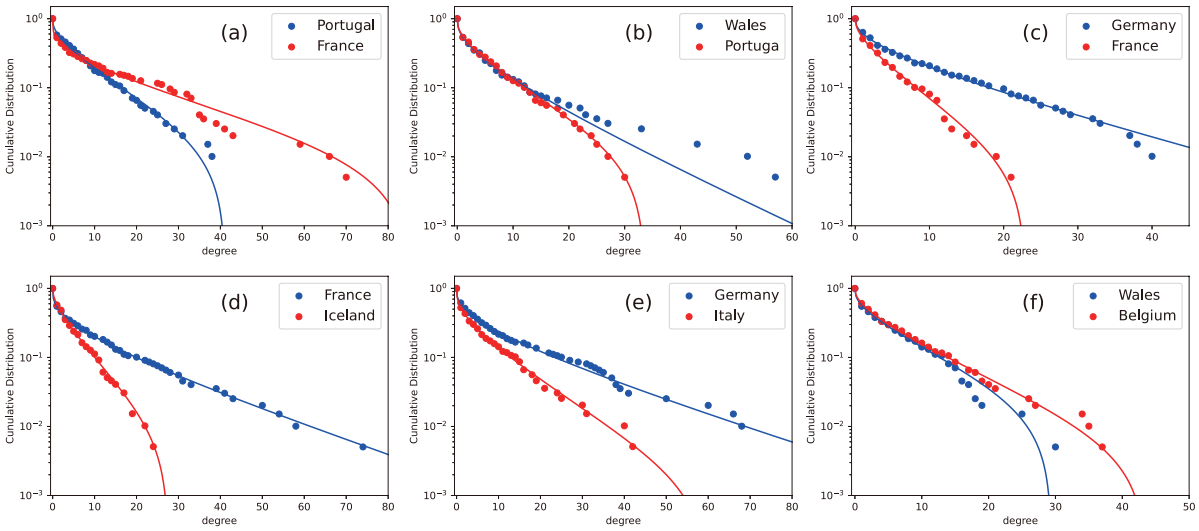


図 4　パスネットワークの次数の累積分布．横軸が次数 k，縦軸が次数の累積確率 $F(k)$ であり，各点が実データであり実線が不完全ガンマ関数での近似曲線である．

表 4　各試合の次数分布を不完全ガンマ関数でフィッティングしたときのパラメータ

試合	チーム	ν	λ	k_{mak}
(a)	ポルトガル	0.40	16.15	41.43
	フランス	0.23	38.07	85.62
(b)	ポルトガル	0.37	12.20	33.82
	ウェールズ	0.34	12.67	298.70
(c)	ドイツ	0.37	17.134	430.59
	フランス	0.43	7.29	23.02
(d)	フランス	0.28	24.27	153.74
	アイスランド	0.48	8.10	27.70
(e)	ドイツ	0.31	25.69	964.45
	イタリア	0.32	13.86	60.07
(f)	ウェールズ	0.39	13.49	29.71
	ベルギー	0.42	12.19	43.69

続いて，重複パスネットワークの累積エッジ多重度分布を図 5 に示す．各分布はほぼ指数関数的に減少している．その結果，多くのエッジが単一エッジを有しており，多重度の高いエッジが少数存在する．特に，(a) の France，(b) の Wales は多重度の高いノードが存在しており，一部の選手・場所にパスが集中していることが分かる．

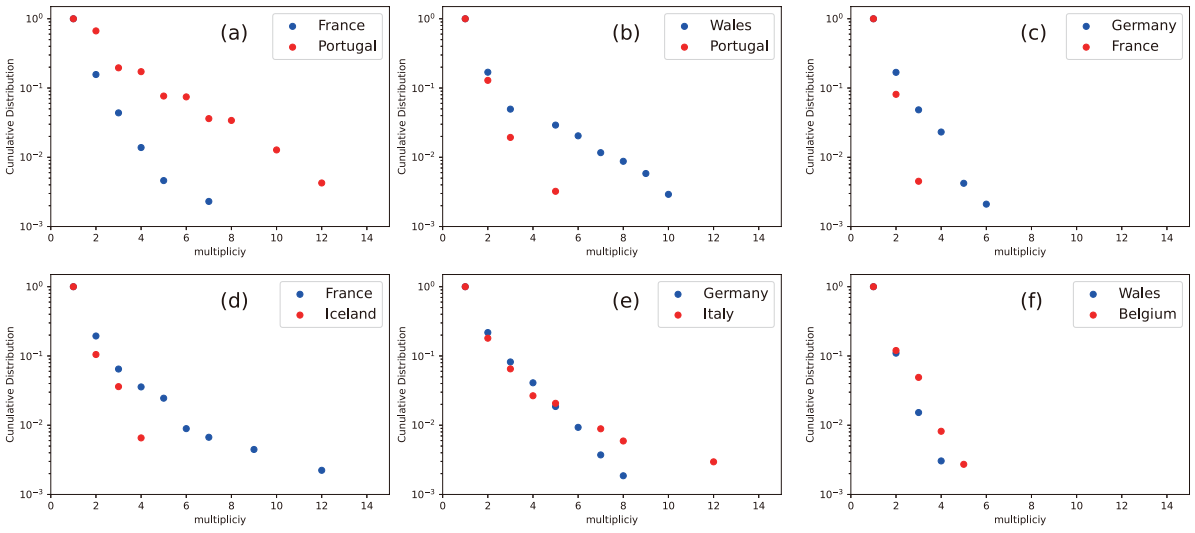


図 5　重複パスネットワークの重複度の累積分布．横軸がエッジの重複度，縦軸がその累積確率である．

5　スペイン 1 部リーグにおけるパスネットワークの次数分布

本節では，前節で得られた統計則の頑健性の検証を行うため，強さが異なるチーム同士の試合におけるパスネットワークの特徴を調べた．前節までの EURO2016 データは 1 試合でも負けたら大会で敗退となるノックアウト式のトーナメントデータであり，偶然の要素が大きくチームの強さの比較が難しい．一方，スペイン一部リーグはリーグ戦であり，9 ヶ月ほどをかけ 38 試合を行うためチームの強さが順位結果に反映されやすい．

そこで，スペイン一部リーグ（2017/2018 シーズン）のリーグ戦最終順位を基準としたチームの強さの違いに着目し，前節と同様にパスネットワークの次数分布について調べた．図 6 は (a) 上位チーム同士の対決，(b) 上位チームと下位チームの対決，(c) 下位チーム同士の対決の試合におけるパスネットワークの次数分布の典型例である．図 6 から分かるようにすべて，不完全ガンマ関数でフィッティングされ，チームの強さの違いによりフィッティング関数は変わらない．ただし，フィッティングパラメータはそれぞれの試合で異なっており表 5 のとおりである．この結果は前節の EURO2016 の結果と一致しており，サッカーにおけるパスネットワークに現れる統計則の頑健性を示唆している．

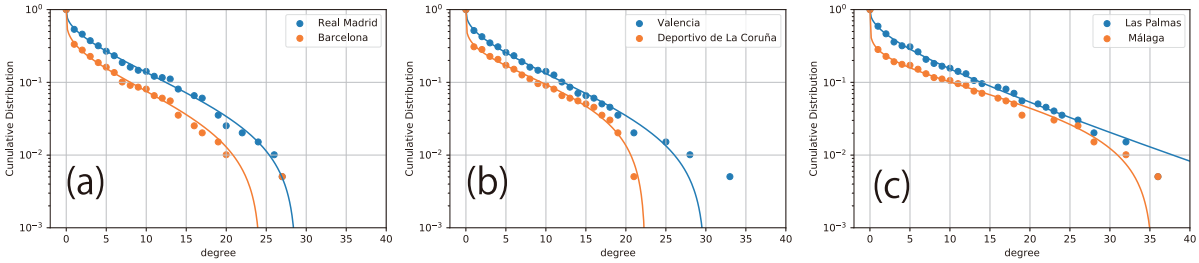


図 6　スペイン一部リーグ 3 試合のパスネットワークの次数の累積分布．横軸が次数 k，縦軸が次数の累積確率 $F(k)$．(a) レアル・マドリード対バルセロナ（上位対決）(b) バレンシア対デポルティーボ（上位・下位対決）(c) ラス・パラモス対マラガ（下位対決）である．各点が実データを表し，実線がフィッティング関数を表す．

表 5　各試合の次数分布を不完全ガンマ関数でフィッティングしたときのパラメータ

試合	チーム	ν	λ	k_{mak}
(a)	バルセロナ	0.18	19.48	24.35
	レアルマドリード	0.34	14.89	28.80
(b)	バレンシア	0.32	15.33	29.97
	デポルティーボ	0.13	150.89	22.46
(c)	ラス・パラモス	0.37	13.03	286.83
	マラガ	0.10	94.80	35.4

6　まとめと今後の課題

本研究では，サッカーの特徴を理解するため先行研究と同様の手法を用い 2016 年 EURO の 6 試合のデータから得られたパスネットワークの統計的特徴を調べた．その結果，6 試合すべてにおいてパスネットワークの次数分布は不完全ガンマ関数でフィッティングできることが分かった．同様に，重複パスネットワークの累積エッジ多重度分布は指数的に減衰することが分かった．どちらの結果とも先行研究の結果 [1] と定性的に一致する．また, スペイン一部リーグのデータを用い，チームの強さの違いに焦点を当てパスネットワークの次数分布を調べた．その結果，EURO2016 の結果と同様に次数分布は不完全ガンマ関数でフィッティングされ，チームの強さの違いに左右されなかった．これらの結果は，パスネットワークの次数分布の特徴は年月によらない頑健な特徴であることを意味している．また，先行研究ではマルコフ連鎖モデルを用いこれらの分布の特徴を再現しており，本研究の結果も踏まえるとサッカーにおけるパスのやり取りは記憶を持たない現象であることが示唆される．

一方，本研究手法は味方同士のパス交換のみを対象としており，ボールを持っていない味方の動きなどは考慮されていない．また，本来サッカーは味方チームと敵チームがお互いに強く干渉しており，同じ位置同じ選手同士のパスであっても敵チームの状況によってその意味は異なる．これらの問題を解決するためにはボールを持たない味方選手の動きや敵選手の動きなど，より詳細なデータの入手が不可欠であり，今後の課題として挙げられる．

本研究の結果は佐藤による 2020 年度武蔵野大学卒業論文 [5] に基づき一部新しい結果を加えたものである．

参考文献

[1] T.Narizuka, K.Yamamoto and Y.Yamazaki, Physica A **412**, 157 (2014).

[2] J.M.Buldú, J. Busquets, J. H. Martínez, J. L. Herrera-Diestra, I. Echegoyen, J. Galeano, J.Luque, Front. Psychol.**9**, 1900 (2018).

[3] L. Pappalardo, P. Cintia, A. Rossi, E. Massucco, P. Ferragina, D. Pedreschi and F. Giannotti, Sci. DATA **6**, 236 (2019).

[4] 片野道郎, データベースの中に世界がある．移籍市場を変えた「Wyscout」, footballista, 2017, https://www.footballista.jp/special/39504, （参照 2021-1-15）.

[5] 佐藤直哉, 2020 年度武蔵野大学工学部数理工学科卒業論文 (2021).

(原稿提出: 2021 年 11 月 19 日; 修正稿提出: 2022 年 01 月 15 日)