Journal of Systemics, Cybernetics and Informatics (2024) 22(5), 48-53
https://doi.org/10.54808/JSCI.22.05.48

An assessment of Markov Chain as a predictive tool for the global performance of an institution of Higher learning

Masikini LUGOMA
Department of Mining, Minerals and Geomatics Engineering
University of South Africa
Pretoria, Gauteng, South Africa

Masengo ILUNGA
Department of Civil and Environmental Engineering and Building Science
University of South Africa
Pretoria, Gauteng, South Africa

Kalenga WAKALENGA
Department of Extraction Metallurgy, University of Johannesburg,
Johannesburg, South Africa

ABSTRACT[1]

The Markov chain (MC) technique is used to make prediction of the ranking of a university in South Africa. The ranking is approached as a stochastic variable. The data were extracted from Scimago in the prediction process. A university was randomly selected to evaluate the level of prediction of the MC method. The university was relatively low ranked. The MC states for this study were adopted from a previous one, based on performance brackets. The Markov chain displayed a satisfactory level of prediction for the probability of a university moving from one level to another. The distribution of university rankings could be well defined to determine the performance of a university.

Keywords: Stochastic process, prediction, ranking, SCImago ranking, performance bracket.

1. INTRODUCTION

The prestige of universities in the global ranking of universities leads to greater admiration in the society and the learning environment. Different ranking systems associated with different methodologies in the ranking process are used. Therefore, there is not a complete agreement about the various criteria or indicators used in the evaluation process. There are different methodologies of university rankings, which could be stable over time, but the rankings vary from one to other [1]. It was emphasized that a tool for monitoring the progress of universities for sustainability in teaching and research, is important [2]. The most known university rankings are the QS World University Rankings**,** Times Higher Education World University Rankings**,** and the Academic Ranking of World Universities (ARWU) (https://www.topuniversities.com/university-rankings-articles/world-university-rankings/world-university-ranking-methodologies-compared). Besides these 3 rankings, the SCImago Institutions Rankings (SIR) is a web-system, considered to be a rating agency that combines three different indicators, namely based on performance: research (50%), innovation results (30%) and social impact (20 %) [3]. The visualisation aspect offered by SCImago is very attractive and user-friendly. Criteria for ranking vary form one ranking to the other.

For instance, the QS World University Rankings is structured around 6 criteria, that include Academic reputation (worth 40% of the overall score), Employer reputation (10%), Student-to-faculty ratio (20%), Research citations per faculty member (20%), Proportion of international faculty (5%), and Proportion of international students (5%). The Times Higher Education World University Rankings, considers 5 indicators, i.e. Teaching (worth 30% of the overall score), Research (30%); research citations (30%), International outlook (7.5%) and Industry income (2.5%). ARWU has 5 criteria, i.e. Alumni (10%), Awards (20%), Highly cited researchers (20%), Papers in Nature and Science (20%), Papers indexed (20%) and Capita performance (10%).

Due to its simplicity and user-friendliness, this study will use the SCimago Institutions Rankings Ranking (SIR). The publication of this ranking system happens on annual basis. It is noted that, namely Quacquarelli Symonds (QS), Times Higher Education (THE), Academic Ranking of World Universities (ARWU), Leiden University ranking

[1] The final version of the paper was peer-edited by Dr Israel Dikgwatlhe, who is acknowledged.

and Webometrics ranking were considered as the evaluate universities worldwide [4].

It is emphasised that predicting deterministically the university ranking seems to be almost impossible. This leads to consider ranking as stochastic or probabilistic event. This type of event (ranking) could be appropriately approached and determined using stochastic methods. One of the most popular of these methods is the Markov Chain [5]. It is stressed that university global performance constitutes an instrument for stakeholders to make a comparison in terms of performance of institutions of higher learning in various diverse fields and disciplines [1]. Hence it can be a powerful instrument for guiding university performance and promoting policies.

2. FOUNDATION OF MARKOV CHAIN

In the Markov chain (MC), the main characteristic in a discrete-time domain is that the future state is only a function of the present, not of the past. Whereas the continuous time in a MC is characterized by the MC process. As said earlier, the MC processes are normally associated with random variables known as stochastic events. These events occur with the only characteristic of dependency only on the present state [5]. Transitions in MC refer to the state transformations of the system. Therefore, the transition probabilities are related to these state transformations. Usually, a transition matrix is used to describe a state space in MC, such that the system moves from an initial state to the next state. It is undeniable that MC has been applied in several fields, i.e. in modelling behaviour of user during web site navigation, e.g. [6],[7] and [8].

Statistical aspects of MC

It is re-iterated that the next state of the system is characterize by the discrete stochastic process such that the probability of changing from one state to the other is function of the previous state and the transition probability between states. Hence MC is mainly composed by a set of states and a set of probability k_{qp} of changing from state q to state p. It is possible that a transition from one state to itself may occur since a university may remain in the same position. Therefore

$k_{qp} = P(Xt+1 = p \mid Xt = q)$ for $q, p \in A$, $t = 0,1,2,$ (1)

where the transition matrix of MC is given by

$$T = [k_{qp}] \quad (2)$$

The probability distribution at different times can be calculated using the transition matrix.

Properties of Markov Chain

Markov chain trajectory: The trajectory of a Markov chain is described by a given set of values taken by a random variable. Thus, this set can be X_0, X_1, X_2, X_3, …. showing the path where the event/subject is in different states.

Absorbing and non-absorbing states: An absorbing state of a MC enables a subject to stay in the same state once it enters. The probability of such a state is 100%. A non-absorbing state or transient state is different from the absorbing state. Thus, a subject may reach an absorbing state from a non-absorbing state, but not the opposite.

Ergodic Markov Chain is characterised by the probability of transition being stationary, meaning the state reached by an event happened steadily. Therefore, the state becomes invariant with time. It was possible to use ergodicity property to analyse the long-term behaviour associated with the dynamics of a university ranking as far as its performance on the global scale is concerned [9].

t-step transition probabilities: Based on partition theorem, it can be shown that over two steps, the probability of a subject to make transition from state q to state p, is as follows:

$P(X_2 = p \mid X_0 = q) = (k^2)_{qp}$.

$P(X_{n+t} = p \mid X_0 = q) = (k^t)_{qp}$.

Probability distribution of random variable: A probability distribution is associated with each random variable Xt.

Given that a row vector of probabilities related to $X_0 \sim \beta$, Hence $X_1 \sim \beta T$ and $X_t \sim \beta T^t$ and generalizing,

$$X_{t+1} \sim \beta T^{t+1}$$

Trajectory Probability: Because of the Markov Property, the probability of any trajectory can be found by multiplying together the starting probability and all subsequent single-step probabilities.

Selected properties of MC have been explained clearly in a simplified manner (https://www.stat.auckland.ac.nz/~fewster/325/notes/ch8.pdf).

The literature presents several modifications to the MC technique e.g. [14], [15], however, a few are illustrated below.

Variants of the Markov Chain

Mobility Markov Chain (MMC): An example is an algorithm for next place prediction based on a mobility model of an individual called a n-MMC that keeps track of the n previous locations visited [5].

Action-based Markov chain: A modeling approach which was developed to predict the window of operating behavior in office spaces, in modelling the occupant dynamically with the building [8].

Cellular automata Markov Chain: The cellular automata -Markov (CA-MC) can be used for simulation and prediction [7], [10], [14].

For instance, in the case of land use land cover change (LULC) [15], the CA-MC, a Markovian chain assessment is carried out for a specific period, usual between years, to produce transition area matrices; subsequently, the transitional area maps are produced for LULC [12]. In this perspective, the accuracy of the model can be used to predict future spatial distribution of land use/change.

Monte Carlo Markov Chain (MCMC): Making and evaluating probabilistic forecasts, MCMC has been shown to be powerful [6].

In Bayesian inference, predictive distributions are typically in the form of samples generated via Markov chain Monte Carlo or related algorithms.

3. DATA AND METHODS

Data used

The period length of the data dated from 2009 and 2023 and were obtained from the "Scimago Institutions Rankings" website. The universities in the previous study [9] were considered in this research. Hence, University of Limpopo (UL), for testing purpose of the methodology. The observation was that this institution of higher learning was ranked within the first 500.

Methodological approach

A bracket of 1000 positions in the ranking constituted the state for the implementation of MC and yielded 8 states over15 years. Transitions happened over those 15-years period. Hence, 8 states were considered with the possibility for a university to move between them over a 15-year period.

The different states are described in Table 1:

Table 1: States of the for the University of Limpopo for Markov Chain, based on SIR Scimago global ranking

University of Limpopo (UL)		
Year	Global Ranking	State
2009	3720	A4
2010	3976	A4
2011	4141	A5
2012	4285	A5
2013	4556	A5
2014	4637	A5
2015	4788	A5
2016	4746	A5
2017	4973	A5
2018	5302	A6
2019	4407	A5
2020	5474	A6
2021	6192	A7
2022	6685	A7
2023	7550	A8

As repeated from the previous study, and shown in Table 1, UL navigated between A4 and A8.

Figure 1 below expresses the MC in a visual representation.

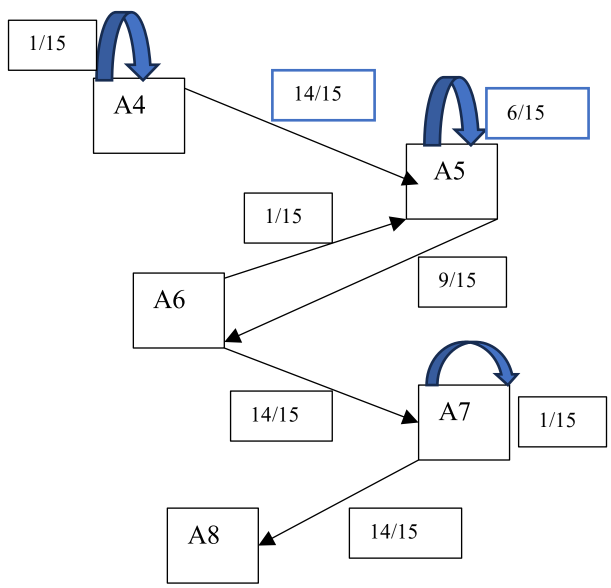

Fig. 1. Markov Chain Configuration for University of Limpopo.

State predictive capability of MC

Given a transition matrix, the prediction of the state of a subject (university) is calculated from its current state vector, considering the university ranking as a random variable (RV) of different states. The distribution of the RV was considered as a function of the different states. Hence, the trajectory probability could be used for the prediction of the university to arrive at a particular state.

For that, different cases were considered arbitrarily with the initial probability of distribution of RV. Therefore, the probability trajectory computation and probability distribution of university ranking were investigated.

The above holds by the property of mutual exclusivity among the states, for a specific university. That implies that a university can only be in one state and not in 2 states, at a given time. The values of the different probabilities in the states were assessed/compared. The transition matrix of UL obtained previously was [9]

$$\begin{bmatrix} 0.067 & 0.933 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0.067 & 0 & 0.933 & 0 \\ 0 & 0 & 0 & 0.067 & 0.933 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

4. RESULTS AND DISCUSSION

Probability trajectory computation

The University of Limpopo probability trajectory derived was explored from Figure 1, as follows:

Given that UL was in state A4, the probability of this university to reach state A5 was determined as follows:

$P(X_2 = A5 | X_0 = A4) =$

$$= \begin{bmatrix} 0.067 & 0.933 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0.933 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0.067 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$= 0.39$

Similarly, the probability for UL to reach the rest the states could be computed as given below:

$P(X_2 = A6 | X_0 = A4) =$

$$= \begin{bmatrix} 0.067 & 0.933 & 0 & 0 & 0 \\ 0 & 0. & 0. & 0 & 0 \\ 0 & 0. & 0 & 0. & 0 \\ 0 & 0 & 0 & 0. & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.067 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$= 0.06$

$P(X_2 = A7 | X_0 = A4) = 0$

$P(X2 = A8 | X_0 = A4) = 0$

$P(X2 = A6 | X_0 = A5) = 0.24$

$P(X2 = A7 | X_0 = A5) = 0.063$

$P(X2 = A8 | X_0 = A5) = 0$

$P(X2 = A7 | X_0 = A6) = 0.06$

$P(X2 = A8 | X_0 = A6) = 0$

$P(X2 = A8 | X_0 = A7) = 0.063$

Generally, the above results showed that the probability values from moving to the next state given that of the previous state was relatively small. Moving from state A4 to A5 was associated with the highest probability (0.4). The future probability of reaching further next stages were most likely to decrease. Given that the start of the ranking is state A4, the most plausible state for UL was the ranking between A4 and A5.

It could be unlikely that most of the time, there will be landing in state A7 or A8, if UL should increase its strategy to research output and relook into its current teaching and learning, and research practices. Hence the least should be state A6 or A5 for this university. If not, the reputation of an institution of higher learning could decrease on the global scale, with the implication of decreasing on the national level. The ranking is not always absolute but associated with a level of perceptions obtained during surveys, which reflect a certain level of subjectivity.

Probability distribution of university ranking as random variable

A given university is equally likely to start in any state at the initial time, supposedly 0.

This implies that the distribution of X_0 is $\beta = (1, 0, 0, 0, 0, 0)$. Therefore, in the case of UL, the distribution of X_0 is $X_1 \sim \beta T$

$$= (1,0,0,0,0) \times \begin{bmatrix} 0.06 & 0.933 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0.067 & 0 & 0.933 & 0 \\ 0 & 0 & 0 & 0.067 & 0.933 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For a given university beginning at state S4 at time 0, the probability distribution of X_2 is such that

$X_2 \sim \beta T^2$

$$= (1,0,0,0,0) \times \begin{bmatrix} 0.06 & 0.933 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0.067 & 0 & 0.933 & 0 \\ 0 & 0 & 0 & 0.067 & 0.933 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} 0.06 & 0.933 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0.067 & 0 & 0.933 & 0 \\ 0 & 0 & 0 & 0.067 & 0.933 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

= (0.004, 0.436, 0.56, 0, 0)

Hence,
P (X_2 = 4) = 0.004, P (X_2 = 5) = 0.436, P (X_3 = 6) = 0.56, P (X_7 = 7) = 0, P (X_2 = 8) = 0.

For UL, this is translated into the probabilities of the university staying in state 4, 7 and 8 as being very negligible, whereas the probability of reaching states 5 and 6 are relatively high and account respectively for 43.6 and 56 %.

A given university is equally likely to start on any state at time 0. The probability of the path trajectory (4, 5, 6, 7, 8) is as follows.

P (4, 5, 6, 7, 8)
= P (X_0 = 4) × p45 × p56 × p67 × p78
= (1/15) × (14/15) × (9/15) × (14/15) × (14/15)
= 0.0325
The case of UL showed that university could reach the final stage with a probability of 3% if the abovementioned path had to be considered. There is still a relatively low chance that UL be trapped into state 8.

5. CONCLUSIONS

This preliminary study focused on the application of the Markov chain method for assessing the Scimago global ranking of a university, namely, the University of Limpopo, which is a South African University. This institution of higher learning is currently in the last state of the global rankings. Deriving predictions of a university ranking motivated this study to inform higher education managers and policy makers towards possible avenues to address the performance of a university. Therefore, the need arises to have a clear knowledge about the future ranking of institutions of higher learning.

Theoretically, Markov Chain demonstrated its capability to predict the future state in terms of the probability trajectory and probability of distribution of ranking. These results revealed that the probability of its ranking to move from the initial state to the next state was close to 40 %, however, moving subsequently to the following level, its probability becomes close to 0. The probability distribution of UL ranking was 0.0325 when the path was from the initial state, thus transiting through all states until the last. The probability distribution of UL ranking, of the next level when the university moved from the initial ranking to the next was (0.004, 0.436, 0.56, 0, 0), with higher probability for states 5 and 6.

In assessing the path taken by university rankings, an understanding of a university ranking, could be drawn.
An assessment of the trajectory probability and the distribution of a university ranking revealed the possible path of the university, hence demonstrating its performance when a university is transiting between different states.

The current study could highlight the predictive capability of MC for the ranking of a university. It could be an opportunity for an institution of higher learning to reposition itself by assessing its strategy for teaching and learning, including research, management, and administrative culture. An improvement plan could be put in place for a university that aspires to move rankings gradually. One could learn from the good practices of universities which have been performing well nationally and worldwide. It will be good to set a benchmark for performance to safeguard the reputation of the university.
Furthermore, the application of the Markov chain technique in evaluating university rankings can contribute to the broader field of higher education management. It provides a quantitative framework for assessing performance trends, making informed decisions, and fostering excellence in academic institutions.

The methodological approach as simplified in this study could be extended to other universities in understanding their dynamics and use other rankings.

6. ACKNOWLEDGMENT

Professor Berhanu Fanta Alemaw (University of Botswana) and Professor John Kabuba Tshilenge (Vaal University of Technology) are acknowledged as peer reviewers of this paper.

7. REFERENCES

[1] F.Selten, C. Neylon, C.K. Huang, & P. Groth, A longitudinal analysis of university rankings. **Quantitative Science Studies**, 1(3), 2020, pp. 1109-1135.

[2] Burmann, C., García, F., Guijarro, F., & Oliver, J. (2021). Ranking the Performance of Universities: The Role of Sustainability. **Sustainability**, 13(23), 13286.

[3] Suprapto, N., Kusnanik, N. W., Iriani, S. S., Wibawa, S. C., Sujarwanto, S., Yulianto, B., ... & Nurhasan, N. (2021). The comparison of Scimago Institutions Rankings (SIR), Scopus, and SINTA profile: A case of the top Indonesian Institutions. **Library Philosophy and Practice (e-journal),** *5788*, 1-11.

[4] Fauzi, M. A., Tan, C. N. L., Daud, M., & Awalludin, M. M. N. (2020). University rankings: A review of methodological flaws. **Issues in Educational Research, 30(1)**, 79-96.

[5] Gambs, S., Killijian, M. O., & del Prado Cortez, M. N. (2012, April). Next place prediction using mobility Markov chains. In **Proceedings of the first workshop on measurement, privacy, and mobility** (pp. 1-6).

[6] Krüger, F., Lerch, S., Thorarinsdottir, T., & Gneiting, T. (2021). Predictive inference based on Markov chain Monte Carlo output. **International Statistical Review, 89(2),** 274-301.

[7] Salakory, M., & Rakuasa, H. (2022). Modeling of Cellular Automata Markov Chain for predicting the carrying capacity of Ambon City. **Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management), 12(2), 372-387.**

[8] Zhou, X., Liu, T., Yan, D., Shi, X., & Jin, X. (2021, April). An action-based Markov chain modeling approach for predicting window operating behavior in office spaces. In **Building Simulation** (Vol. 14, pp. 301-315). Tsinghua University Press.

[9] M. Lugoma, M. Ilunga (2024) Proceedings of the 15th International Conference on Society and Information Technologies (ICSIT 2024), pp.54-58, https://doi.org/10.54808/ICSIT2024.01

[10] Koko, A. F., Yue, W., Abubakar, G. A., Hamed, R., & Alabsi, A. A. N. (2020). Monitoring and predicting spatio-temporal land use/land cover changes in Zaria City, Nigeria, through an integrated cellular automaton and Markov chain model (CA-Markov). **Sustainability**, 12(24), 10452.

[11] Albasri, N. A. H., Al-Jawari, S. M., & Al-Mosherefawi, O. J. (2022). Prediction of urban spatial changes pattern using Markov chain. **Civil Engineering Journal**, 8(4), 710-722.

[12] S.W. Wang, L. Munkhnasan and W.-K. Lee (2021). use and land cover change detection and prediction in Bhutan's high-altitude city of Thimphu, using cellular automata and Markov chain. Environmental Challenges 2, 2021, 100017.

[13] Wang, S. W., Gebru, B. M., Lamchin, M., Kayastha, R. B., & Lee, W. K. (2020). Land use and land cover change detection and prediction in the Kathmandu district of Nepal using remote sensing and GIS. **Sustainability**, 12(9), 3925

[14] Zhu, J., Hong, J., & Hughes, J. G. (2002, April). Using Markov chains for link prediction in adaptive web sites. In **Int. Conference on Soft Issues in the Design, Development, and Operation of Computing Systems** (pp. 60-73). Berlin, Heidelberg: Springer Berlin Heidelberg.

[15] Mallak, Saed; Kanan, Mohammad; Al-Ramahi, Nidal; Qedan, Aya; Khalilia, Hadi; and Khassati, Ahmad (2023) "Using Markov Chains and Data Mining Techniques to Predict Students' Academic Performance," **Information Sciences Letters**: Vol. 12: Iss. 9, PP -. Available at: https://digitalcommons.aaru.edu.jo/isl/vol12/iss9/15